UNIVERSITÀ DEGLI STUDI DI ROMA
TOR VERGATA

Dottorato in Scienze Forensi
XVII Ciclo

Coordinatore: Professor Giovanni Arcudi

MATHEMATICAL APPROACH IN FORENSIC SCIENCE AND USE OF PROBABILITY IN EVALUATION OF EVIDENCE

APPROCCIO MATEMATICO NELLE SCIENZE FORENSI ED USO DELLA PROBABILITÀ NELLA VALUTAZIONE DELLA TRACCIA

Tesi di Dottorato del
dottor Bruno CARDINETTI

Anno 2004

PREFACE

- *And now, I said, let me show in a figure how far our nature is enlightened or unenlightened: Behold! human beings living in an underground den, which has a mouth open toward the light and reaching all along the den; here they have been from their childhood, and have their legs and necks chained so that they cannot move, and can only see before them, being prevented by the chains from turning round their heads. Above and behind them a fire is blazing at a distance, and between the fire and the prisoners there is a raised way; and you will see, if you look, a low wall built along the way, like the screen which marionette players have in front of them, over which they show the puppets.*
- *I see.*
- *And do you see, I said, men passing along the wall carrying all sorts of vessels, and statues and figures of animals made of wood and stone and various materials, which appear over the wall? Some of them are talking, others silent.*
- *You have shown me a strange image, and they are strange prisoners.*
- *Like ourselves, I replied; and they see only their own shadows, or the shadows of one another, which the fire throws on the opposite wall of the cave?*
- *True, he said; how could they see anything but the shadows if they were never allowed to move their heads?*
- *And of the objects which are being carried in like manner they would only see the shadows?*
- *Yes, he said.*
- *And if they were able to converse with one another, would they not suppose that they were naming what was actually before them?*
- *Very true.*
- *And suppose further that the prison had an echo which came from the other side, would they not be sure to fancy when one of the passersby spoke that the voice which they heard came from the passing shadow?*
- *No question, he replied.*
- *To them, I said, the truth would be literally nothing but the shadows of the images.*
- *That is certain.*

Plato, *Republic*, VII, 514-515 (translation by Benjamin Jowett)

Like in Plato's myth of cave, the judge in the Court can be considered as a prisoner who is trying to know objects by seeing their shadows. Actually, in legal proceedings only the "shadows" of the crime can be visible, where the "shadows" are evidences, witnessing,

documents, etc.
And in his grave duty of decision making, the judge often involves other prisoners, and in particular the "forensic scientist", who has to supply the judge in evaluation of evidences. In order to give the same meaning to the word "forensic scientist", in the last years particular attention has been focused on the educational background that people involved in forensic science should have. In this direction, universities and other associations have more and more concentrated their effort in creating a well-defined professional skill.
In this sense, the author acknowledges the Università degli Studi "Tor Vergata" in Rome - in particular Professor Giovanni Arcudi - and his colleagues of the Raggruppamento Carabinieri Investigazioni Scientifiche for the daily discussions in forensic science - Captain Davide Zavattaro, in particular.
Special thanks to Professor Camillo Cammarota of the Mathematical Department of Università degli Studi "La Sapienza" in Rome for the important contribution to this work.

CONTENTS

1

INTRODUCTION

Fundamental principle of Criminal Law is that, when a crime is committed, the competent Prosecutor Office and Court are arranged beforehand on the basis of the type of the crime, the place where it has been committed and the condition of the author of the crime.

Referring in particular to Italian Criminal Law, when a *notitia criminis* is known, two subjects of legal proceedings are necessary: the prosecutor, and the judge (or jury). Respectively, they have to propose and make decision in order to reach a filing of the case, or a formal accusation against one or more people: in this last case, besides the prosecutor and the judge, also the accused and his defender become subjects of legal proceedings.

Moreover, other subjects can act legal proceedings (the plaintiff, civil party, etc.). In particular the judicial police have the duty to support the prosecutor in order to acquire the *notitia criminis*, to avoid that crimes have further consequences, to search the authors, to do all necessary in assuring and collecting the sources of proofs. In this framework, the judicial police operate in the scene of crime in order to collect exhibits and evidences, where the term of "scene of crime" is to be considered in an extended sense: not only as the physical place where the crime has been committed, but also as everything has had a connection to the crime. In fact, collection of evidences on a person (e.g. gunshot residues on the hands of a suspect), or on a virtual space (e.g. files on a web site) can be considered as acts on a scene of crime.

When the formal process starts, the aim of the judge in Criminal Court is decision making about the subsistence of the crime, the responsibility of the accused, and the other parameters in order to determine the penalty, the application of precautionary measures and the responsibility of the accused according to the Civil Law.

In decision making on these features, the prosecutor at first, together with the judicial police, and then the judge (who is the *peritus peritorum*, i.e "the expert among the experts"), have to discriminate two opposite hypotheses. For convention, the opposite hypotheses are called *prosecutor's* and *defender's*, respectively, even if they are not necessarily asserted by this two subjects.

Note that the judge has the moral duty in always researching the truth, and, in doubt, even if the prosecutor and the defender agree in the version of the affair, he must more closely investigate the matter (e.g. the prosecutor and the defender could agree that the accused killed a police officer in a shooting, but the doubt that another policeman fired the colleague could arise).

The discrimination of the two opposite hypotheses must be evaluated on the basis of proofs, which are proposed to the judge during the debate by the prosecutor and the defender by means of different means of proof: witnessing, examination of the parties, cross-examination, line-up, judicial test, expertise, and documents.

Obviously, the discrimination of the two opposite hypotheses by the evaluation of the proofs could not be trivial, and methods to quantify it are useful.

In this setting, the aim of forensic science is to support the judge in decision making in Court by means of a scientific approach in the evaluation of evidences.

Historically, different branches of forensic science have developed proper ways in analyzing and evaluating evidences, trying to answer to questions like the following: Are two fingerprints produced by the same finger? Are two bullets shot by the same firearm? Are two fragments of glass from the same window? Is this substance heroine? Is the recorded sound the suspect's voice? etc.

In order to answer to these questions, a process of comparison between two objects (e.g. two bullets, two fingerprint, two fragments of glass, etc.), or an object and a standard (e.g. the substance and the standard of what is legally defined as heroine), is required.

In the last years, the forensic scientific community (American Academy of Forensic Science, AAFS; European Academy of Forensic Science, EAFS; European Network of Forensic Science Institutes, ENFSI; Interpol; etc.) has proposed the use of scientific methods for the evaluation of evidences, based on probability. The Bayesian approach subsequently clarifies and restrains the role of the forensic scientist in legal proceedings, and forces him to present the results of his analyses in a such appropriate way that no doubt of interpretation should arise. Hence, a "beyond a reasonable doubt" truth should be overcome by a quantification of an error rate, a confidence level, a reliability level, a likelihood ratio, etc.

Questions about the admissibility of the scientific evidence in Court are much-discussed. In particular, objections to the use of probability in legal proceedings are based on the

presumed breach of the causality principle, according to which the proof of the personal responsibility requires that the occurred behaviour be a necessary condition for the crime.
This thesis proposes a general panorama of the use of mathematics in forensic science, trying to give a systematic approach to the general principles of identity, and to the processes of identification and individualization. Moreover, the probability approach in evaluation of evidences is introduced, and its applications in particular non-standard forensic branches are proposed.
It is author's belief that the language of science (and then of forensic science) is only mathematics. Although its rigor, mathematics has practical use in daily forensic scientist job, and this is the reason for which a formal mathematical approach is here suggested.

REFERENCES

[1]. Tuthill Harold, *Individualization: principles and procedures in criminalistics*, Lightning Powder Company Inc. (1994)

[2]. *Codice di Procedura Penale* (2004)

[3]. Quintili Michela, Cappelleri Domenico, Coppola Thomas, *La prova particolaristica della "causalità individuale"*, from Stella Federico, *Giustizia e modernità* (2001)

[4]. Silvia Aquitani, Alberto Buldini, *La mancanza di certezze della scienza*, from Stella Federico, *Giustizia e modernità* (2001)

[5]. Cooley Craig M., *Daubert and Kumho Tire: a wake up call for the forensic sciences*, from `http://www.law-forensic.com/library.htm` (2004)

SECTION I

GENERALITIES

2

BASIC MATHEMATICAL CONCEPTS

In order to proceed to a formalization of the fundamental principles of forensic science, basic mathematical notations and results are here summarized, without any proposal of completeness. Demonstrations of theorems are omitted if retained trivial or normally reported in standard mathematical literature.

2.1. SET THEORY

Let $X \neq \emptyset$ and $Y \neq \emptyset$ denote in general two non-empty sets.

Definition 2.1.1. If $A \subseteq X$, the *power set of* A, $\wp A$, denotes the collection of all subsets of A:

$$\wp A = \{B \subseteq X | B \subseteq A\} \quad .$$

Note 2.1.1. Obviously, $\wp X$ represents the collection of all subsets of X **.**

Definition 2.1.2. Let $J \neq \emptyset$ denote an index set, and let $\{X_j\}_{j \in J}$ be a family of non-empty sets. Then, $\forall K \subseteq J$, $K \neq \emptyset$, the *(Cartesian) product set* X_K, indicated by:

$$X_K = \prod_{j \in K} X_j$$

is the set whose elements $x_K \in X_K$ are the ordered sequences $x_K = (x_j)_{j \in K}$, with $x_j \in X_j$ for all $j \in K$.

Theorem 2.1.1. Let $J \neq \emptyset$ denote an index set, and let $\{X_j\}_{j \in J}$ be a family of non-empty sets. Then, $\forall K \subseteq J$, $K \neq \emptyset$, the set X_J can be written as:

$$X_J = X_K \times X_{J-K}$$

The element $x_J \in X_J$ can be written as $x_J = (x_K, x_{J-K})$, with $x_K \in X_K$ and $x_{J-K} \in X_{J-K}$.

2.2. RELATIONS

Definition 2.2.1. A *unitary relation Q on a set X*, also called *property*, is a subset Q of X, i.e. $Q \in \wp X$.

Definition 2.2.2. A *binary relation R on a set X and a set Y* is a subset of the Cartesian product of X and Y, i.e. $R \in \wp(X \times Y)$. Note that, $\forall x \in X$ and $\forall y \in Y$, an alternative notation of $(x, y) \in R$ is xRy .

Definition 2.2.3. A *partial order relation O on a set X* is a binary relation on X^2 such that, $\forall x_1, x_2, x_3 \in X$:
(a). *reflexivity*: $x_1 O x_1$;
(b). *antisymmetry*: if $x_1 O x_2$ and $x_2 O x_1$, then $x_1 = x_2$;
(c). *transitivity*: if $x_1 O x_2$ and $x_2 O x_3$, then $x_1 O x_3$.

Definition 2.2.4. Let O a partial order relation on a set X. The pair (X, O) is called a *partial ordered set* .

Definition 2.2.5. On a partial ordered set (X, O) an irreflexive relation O' on the set X is defined, $\forall x_1, x_2 \in X$, by:

$$x_1 O' x_2 \quad , \quad \text{if } x_1 O x_2 \text{ and } x_1 \neq x_2 \quad .$$

Definition 2.2.6. An *equivalence relation D on a set X* is a binary relation on X^2 such that, $\forall x_1, x_2, x_3 \in X$:
(a). *reflexivity*: $x_1 D x_1$;
(b). *symmetry*: if $x_1 D x_2$, then $x_2 D x_1$;
(c). *transitivity*: if $x_1 D x_2$ and $x_2 D x_3$, then $x_1 D x_3$.

Example 2.2.1. On a set X, many different equivalence relations can be defined. In particular, the *concrete* or *banal equivalence relation* for which $\Delta = \{(x, x) \in X^2 | x \in X\}$, and the *discrete equivalence relation* for which $D = X^2$.

Theorem 2.2.1. Let $J \neq \emptyset$ denote a finite index set, and let $\{D_j\}_{j \in J}$ be a family of equivalence relations on a set X. Then, the intersection $\bigcap_{j \in J} D_j$ is an equivalence relation on X .

Example 2.2.2. Let F be a binary relation on X^2, i.e. $F \subseteq \wp X^2$. The set of all equivalence relations containing F is not empty, and the intersection of all these equivalence relations is the smallest equivalence relation containing F, called the *equivalence relation generated by* F and denoted by $D(F)$. Note that $\forall F \subseteq \wp X^2$, $\Delta \subseteq D(F) \subseteq \wp X^2$, i.e. the concrete and the discrete equivalence relations are the smallest and the largest ones, respectively .

Theorem 2.2.2. Let Q be a property on a set X. On the set X, a equivalence relation D_Q is naturally defined, such that $\forall (x_1, x_2) \in X^2$:

$$x_1 D_Q x_2 \quad , \quad \text{if } x_1 \in Q \text{ and } x_2 \in Q \quad .$$

Definition 2.2.7. Let D be an equivalence relation on a set X. Then, $\forall x \in X$, the *equivalence class* $[x]_D$ *of* x is defined by:

$$[x]_D = \{y \in X | xDy\} \quad .$$

Definition 2.2.8. The set of all equivalence classes defined by the equivalence relation D on a set X is called *quotient set* and denoted by X/D .

Theorem 2.2.3. Let D be an equivalence relation on a set X. Then:

$$\bigcup_{x \in X/D} [x]_D = X \quad .$$

Theorem 2.2.4. Let D be an equivalent relation on a set X. On the set X, $\forall x \in X$, a family of properties $Q_{x,D}$ is naturally defined, such that $\forall y \in X$:

$$y \in Q_{x,D} \quad , \quad \text{if } y \in [x]_D$$

The elements y is called to have the same D-property of x .

2.3. FUNCTIONS

Definition 2.3.1. A *function* f *on a set* X *into a set* Y is binary relation on $X \times Y$ such that, $\forall x_1, x_2 \in X$ and $\forall y_1, y_2 \in Y$:

(a). *univocity*: if $(x_1, y_1) \in f$ and $(x_2, y_2) \in f$, then $x_1 = x_2$ implies $y_1 = y_2$.

Note that an alternative notation of the function $f \in \wp(X,Y)$ is $f : X \to Y$, and, $\forall x \in X$ and $\forall y \in Y$, an alternative notation of $(x,y) \in f$ is $y = fx$.

Definition 2.3.2. If $f : X \to Y$ is a function, the *power map of the function* f, $f : \wp X \to \wp Y$, is defined, $\forall A \in \wp X$, by:

$$fA = \{y \in Y | \exists x \in A : y = fx\} \quad .$$

Definition 2.3.3. If $f : X \to Y$ is a function, the *complete inverse image map* f^{-1} *of the function* f, $f^{-1} : \wp Y \to \wp X$, is defined, $\forall B \in \wp Y$, by:

$$f^{-1}B = \{x \in X | fx \in B\}$$

In particular, let $f^{-1}x$ denote $f^{-1}\{x\}$.

Theorem 2.3.1. If $f : X \to Y$ is a function, $\forall A_1, A_2 \in \wp X$ and $\forall B_1, B_2 \in \wp Y$:
(a1). $A_1 \subseteq f^{-1}(fA_1)$;
(a2). $f(f^{-1}B_1) \subseteq B_1$;
(b1). $f(A_1 \cap A_2) \subseteq (fA_1) \cap (fA_2)$;
(b2). $f(A_1 \cup A_2) = (fA_1) \cup (fA_2)$;
(c1). $f^{-1}(B_1 \cap B_2) = (f^{-1}B_1) \cap (f^{-1}B_2)$;
(c2). $f^{-1}(B_1 \cup B_2) = (f^{-1}B_1) \cup (f^{-1}B_2)$.

Definition 2.3.4. If $A \in \wp X$, the *characteristic function* Φ_A, $\Phi_A : X \to \{0,1\}$, is defined, $\forall x \in X$, by:

$$\Phi_A x = \begin{cases} 1 & , \quad \text{if } x \in A \\ 0 & , \quad \text{if } x \notin A \end{cases} \quad .$$

Definition 2.3.5. Let $J \neq \emptyset$ denote a finite index set, and let $\{A_j\}_{j \in J}$ be a family a subset of X. If $\{\alpha_j\}_{j \in J}$ is a family of real numbers, then the finite linear combination of characteristic functions $\phi : X \to R$ defined by:

$$\phi = \sum_{j \in J} \alpha_j \Phi_{A_j}$$

is called *simple function* .

Definition 2.3.6. A simple function ϕ is called *positive* if $\phi : X \to R^+$.

Definition 2.3.7. Let f be a real function, i.e. $f : X \to R$. Then:
(a). $f^+ = \sup(0, f)$ is called the *positive part of the function* f ,
(b). $f^- = \sup(0, -f)$ is called the *negative part of the function* f .

Definition 2.3.8. Let $J \neq \emptyset$ denote an index set, and let $\{X_j\}_{j \in J}$ be a family of non-empty sets. Then, $\forall K \subseteq J$, $K \neq \emptyset$, the *projection map* $p^J_K : X_J \to X_K$ is the function defined, $\forall x_J \in X_J$ with $x_J = (x_K, x_{J-K})$, by:

$$p^J_K x_J = x_K$$

For simplicity, $\forall j \in J$, let p_j indicate the projection map $p^J_{\{j\}}$.

Note 2.3.1. Let D be an equivalence relation on a set X. Then the function $f : X \to X/D$ defined, $\forall x \in X$, by:

$$fx = [x]_D$$

is naturally defined .

Note 2.3.2. Given a function $f : X \to Y$, an equivalence relation D on a set X is naturally defined, $\forall x_1, x_2 \in X$, by:

$$x_1 D x_2 \quad , \quad \text{if } fx_1 = fx_2 \quad .$$

2.4. TOPOLOGY

Definition 2.4.1. A *topology* $\mathcal{T}$ *on a set* X is a subset of $\wp X$, $\mathcal{T} \subseteq \wp X$, such that:
(a). $\{\emptyset, X\} \subseteq \mathcal{T}$;
(b). if $\{G_1, G_2\} \subseteq \mathcal{T}$, then $G_1 \cap G_2 \in \mathcal{T}$;
(c). if $\{G_j\}_{j \in J} \subseteq \mathcal{T}$, then $\bigcup_{j \in J} G_j \in \mathcal{T}$.

Definition 2.4.2. The pair $(X, \mathcal{T})$ is called a *topological space* .

Example 2.4.1. On a set X, many different topologies can be defined. In particular, the *concrete* or *banal topology* for which $\mathcal{T} = \{\emptyset, X\}$ and the *discrete topology* for which $\mathcal{T} = \wp X$.

Theorem 2.4.1. Let $J \neq \emptyset$ denote a finite index set, and let $\{\mathcal{T}_j\}_{j \in J}$ be a family of topologies on a set X. Then, the intersection $\bigcap_{j \in J} \mathcal{T}_j$ is a topology on X .

Example 2.4.2. Let $\mathcal{F}$ be a subset of $\wp X$, i.e. $\mathcal{F} \subseteq \wp X$. The set of all topologies containing $\mathcal{F}$ is not empty, and the intersection of all these topologies is the smallest topology containing $\mathcal{F}$, called the *topology generated by* $\mathcal{F}$ and denoted by $\mathcal{T}(\mathcal{F})$. Note that $\forall \mathcal{F} \subseteq \wp X$, $\{\emptyset, X\} \subseteq \mathcal{T}(\mathcal{F}) \subseteq \wp X$, i.e. the concrete and the discrete topologies are the smallest and the largest ones, respectively. Finally, $\forall G \in \wp X$, $\mathcal{T}(\{G\}) = \{\emptyset, G, X\}$.

Definition 2.4.3. Let $(X, \mathcal{T})$ be a topological space. The element $G \in \mathcal{T}$ is called *open set* .

Definition 2.4.4. Let $(X,\mathcal{T})$ be a topological space. The element $H \in \wp X$ is called *closed set* if $\exists G \in \mathcal{T}$ such that $H = X - G$.

Definition 2.4.5. Let $(X,\mathcal{T})$ be a topological space and $A \in \wp X$. The *interior of* A, A^0, is the union of all open subsets of A:

$$A^0 = \bigcup\{G \in \wp X | G \subseteq A, G \in \mathcal{T}\}$$

and it is the largest open subset of A .

Theorem 2.4.2. Let $(X,\mathcal{T})$ be a topological space. Then, $\forall A \in \wp X$, $A^0 \subseteq A \subseteq A^C$.

Definition 2.4.6. Let $(X,\mathcal{T})$ be a topological space and $A \in \wp X$. The *closure of* A, A^-, is the intersection of all closed sets containing A:

$$A^- = \bigcap\{H \in \wp X | A \subseteq H, X - H \in \mathcal{T}\}$$

and it is the smallest closed set containing A .

Definition 2.4.7. Let $(X,\mathcal{T})$ be a topological space and $A \in \wp X$. The *boundary* or *frontier* or *derived set of* A, ∂A, is defined by:

$$\partial A = A^- - A^0$$

and it is the smallest closed set containing A .

Definition 2.4.8. Let $(X,\mathcal{T})$ be a topological space and $x \in X$. A set $A \in \wp X$ is a *neighborhood of* x if $x \in A^0$.

Definition 2.4.9. Let $(X,\mathcal{T})$ and $(Y,\mathcal{S})$ be two topological spaces. A function $f : X \to Y$ is called $\mathcal{T}$-$\mathcal{S}$-*continuous* if $\forall G \in \mathcal{S}$, $f^{-1}G \in \mathcal{T}$.

Theorem 2.4.3. Let $(X,\mathcal{T})$ be a topological space. The function $f : X \to Y$ naturally defines a topology $\mathcal{S}$ on Y:

$$\mathcal{S} = \{B \subseteq Y | f^{-1}B \in \mathcal{T}\} \quad .$$

Theorem 2.4.4. Let $(Y,\mathcal{S})$ be a topological space. The function $f : X \to Y$ naturally defines a topology $\mathcal{T}$ on X:

$$\mathcal{T} = \{A \subseteq X | A = f^{-1}B, B \in \mathcal{S}\} \quad .$$

Definition 2.4.10. Let $<$ be the irreflexive relation on a partial ordered set $(X,\leq)$. The *order topology on* X is the topology generate by the intervals of the type $B_a^- = \{x \in X | x < a\}$ and $B_a^+ = \{x \in X | a < x\}$, $\forall a \in X$.

Definition 2.4.11. The *Euclidean topology on R* is the order topology generate by the intervals of the type $(-\infty, a)$ and $(a, +\infty)$, $\forall a \in R$.

Definition 2.4.12. Let $J \neq \emptyset$ denote an index set, and let $\{(X_j, \mathcal{T}_j)\}_{j\in J}$ be a family of topological spaces. Then, $\forall K \subseteq J$, $K \neq \emptyset$, the *product topology* $\mathcal{T}_K = \bigotimes_{j\in K} \mathcal{T}_j$ on the product set X_K is the topology generated by the subsets $A_K = \prod_{j\in K} A_j$ with $A_j \in \mathcal{T}_j$, $\forall j \in K$.

Definition 2.4.13. Let $J \neq \emptyset$ denote an index set, and let $\{(X_j, \mathcal{T}_j)\}_{j\in J}$ be a family of topological spaces. Then, $\forall K \subseteq J$, $K \neq \emptyset$, the pair $(X_K, \mathcal{T}_K)$ is called the *product of the topological spaces* $\{(X_j, \mathcal{T}_j)\}_{j\in J}$, and indicated by:

$$\bigotimes_{j\in K} (X_j, \mathcal{T}_j) \quad .$$

Theorem 2.4.5. Let $J \neq \emptyset$ denote an index set, and let $(X_j, \mathcal{T}_j)$, $j \in J$, be a family of topological spaces. Then, $\forall K \subseteq J$, the projection map p^J_K is $\mathcal{T}_J$-$\mathcal{T}_K$-continuous .

2.5. σ-ALGEBRA

Definition 2.5.1. A *σ-algebra $\mathcal{A}$ on a set X* is a subset of $\wp X$, $\mathcal{A} \subseteq \wp X$, such that:
(a). $X \in \mathcal{A}$;
(b). if $A \in \mathcal{A}$, then $X - A \in \mathcal{A}$;
(c). if $\{A_j\}_{j\in J} \subseteq \mathcal{A}$, then $\bigcup_{j\in J} A_j \in \mathcal{A}$.

Definition 2.5.2. The pair $(X, \mathcal{A})$ is called a *measurable space* .

Example 2.5.1. On a set X, many different σ-algebras can be defined. In particular, the *concrete* or *banal σ-algebra* for which $\mathcal{A} = \{\emptyset, X\}$ and the *discrete σ-algebra* for which $\mathcal{A} = \wp X$.

Theorem 2.5.1. Let $J \neq \emptyset$ denote a finite index set, and let $\{\mathcal{A}_j\}_{j\in J}$ be a family of σ-algebras on a set X. Then, the intersection $\bigcap_{j\in J} \mathcal{A}_j$ is a σ-algebra on X .

Example 2.5.2. Let $\mathcal{F}$ be a subset of $\wp X$, i.e. $\mathcal{F} \subseteq \wp X$. The set of all σ-algebras containing $\mathcal{F}$ is not empty, and the intersection of all these σ-algebras is the smallest σ-algebra containing $\mathcal{F}$, called the *σ-algebra generated by $\mathcal{F}$* and denoted by $\mathcal{S}(\mathcal{F})$. Note that $\forall \mathcal{F} \subseteq \wp X$, $\{\emptyset, X\} \subseteq \mathcal{S}(\mathcal{F}) \subseteq \wp X$, i.e. the concrete and the discrete σ-algebras are the smallest and the largest ones, respectively. Finally, $\forall A \in \wp X$, the σ-algebra generated by $\{A\}$ is $\mathcal{S}(\{A\}) = \{\emptyset, A, X - A, X\}$.

Definition 2.5.3. Let $(X, \mathcal{A})$ be a measurable space. The element $A \in \mathcal{A}$ is called *measurable set* .

Definition 2.5.4. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be two measurable spaces. A function $f : X \to Y$ is called *$\mathcal{A}$-$\mathcal{B}$-measurable* if $\forall B \in \mathcal{B}$, $f^{-1}B \in \mathcal{A}$.

Theorem 2.5.2. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be two measurable sets. If $f : X \to Y$ is constant, the the function f is $\mathcal{A}$-$\mathcal{B}$-measurable .

Theorem 2.5.3. Let $(X, \mathcal{A})$ be a measurable space. The function $f : X \to Y$ naturally defines a σ-algebra $\mathcal{B}$ on Y:

$$\mathcal{B} = \{B \subseteq Y | f^{-1}B \in \mathcal{A}\} \quad .$$

Theorem 2.5.4. Let $(Y, \mathcal{B})$ be a measurable space. The function $f : X \to Y$ naturally defines a σ-algebra $\mathcal{A}$ on X:

$$\mathcal{A} = \{A \subseteq X | A = f^{-1}B, B \in \mathcal{B}\} \quad .$$

Definition 2.5.5. Let $(X, \mathcal{T})$ be a topological space. The σ-algebra generated by the open sets of X, $B = \mathcal{S}(\mathcal{T})$, is called *Borel σ-algebra*, and its elements are called *Borel sets*. The pair $(X, \mathcal{B})$ is called *Borel measurable space* .

Definition 2.5.6. Let $(X, \mathcal{A})$ be a measurable set and let $(Y, \mathcal{B})$ be a Borel measurable spaces. A $\mathcal{A}$-$\mathcal{B}$-measurable function $f : X \to Y$ is called *Borel measurable* .

Definition 2.5.7. The Borel σ-algebra B on R is the σ-algebra generated by the open space of the Euclidean topology on R. The measurable space (R, B) is called *real Borel measurable space* .

Theorem 2.5.5. Let $(X, \mathcal{A})$ be a measurable set and let (R, B) be a real Borel measurable spaces. If $f : X \to R$ and $g : X \to R$ are two Borel measurable functions, then:
(a). $\forall a \in R^+$, $|f|^a$ is Borel measurable ;
(b). if $f \neq 0$, $\frac{1}{f}$ is Borel measurable ;
(c). $f + g$ is Borel measurable ;
(d). $f \cdot g$ is Borel measurable .

Definition 2.5.8. Let $J \neq \emptyset$ denote an index set, and let $\{(X_j, \mathcal{A}_j)\}_{j \in J}$ be a family of measurable spaces. Then, $\forall K \subseteq J$, $K \neq \emptyset$, the *product σ-algebra* $\mathcal{A}_K = \bigotimes_{j \in K} \mathcal{A}_j$ on the product set X_K is the σ-algebra generated by the subsets $A_K = \prod_{j \in K} A_j$ with $A_j \in \mathcal{A}_j$, $\forall j \in K$.

Definition 2.5.9. Let $J \neq \emptyset$ denote an index set, and let $\{(X_j, \mathcal{A}_j)\}_{j \in J}$ be a family of measurable spaces. Then, $\forall K \subseteq J$, $K \neq \emptyset$, the pair $(X_K, \mathcal{A}_K)$ is called the *product of the measurable spaces* $\{(X_j, \mathcal{A}_j)\}_{j \in J}$, and indicated by:

$$\bigotimes_{j \in K} (X_j, \mathcal{A}_j) \quad .$$

Theorem 2.5.6. Let $J \neq \emptyset$ denote an index set, and let $\{(X_j, \mathcal{A}_j)\}_{j \in J}$ be a family of measurable spaces. Then, $\forall K \subseteq J$, the projection map p_K^J is $\mathcal{A}_J$-$\mathcal{A}_K$-measurable .

2.6. THEORY OF MEASURE

Definition 2.6.1. Let $(X, \mathcal{A})$ be a measurable space. A *(positive) measure* μ *on the measurable space* $(X, \mathcal{A})$ is a function $\mu : \mathcal{A} \to R^+ \cup \{+\infty\}$ such that:
(a). $\mu(\emptyset) = 0$;
(b). *numerable additivity*: if $\{A_j\}_{j \in J} \subseteq \mathcal{A}$, with $\{A_j\}_{j \in J}$ disjoint sets, then:

$$\mu(\bigcup_{j \in J} A_j) = \sum_{j \in J} \mu(A_j) \quad .$$

Definition 2.6.2. The triade $(X, \mathcal{A}, \mu)$ is called *measure space* .

Example 2.6.1. Let $(X, \wp X)$ be a measurable space. If $A \in \wp X$, the *counting measure* $\sharp$ *of the set* A, $\sharp A$, is defined by the number of points in A, if A is finite; and by $+\infty$ if A is infinite .

Example 2.6.2. *Lebesgue measure.* Let (R, B) be the Borel real measurable space, and $A \in B$. Denoting by $\ell(I)$ the length of an open interval $I \subseteq R$, the *outer measure of* A is defined by:

$$m^* A = \inf \sum_{j \in J} \ell(I_j)$$

for all $\{I_j\}_{j \in J}$ such that $A \subseteq \bigcap_{j \in J} I_j$. The *inner measure of* A is defined by:

$$m_* A = \sup_{K \subseteq A} m^* K$$

for all closed set $K \subseteq A$. The subset A is *Lebesgue measurable* if $m_* A = m^* A$, and the measure is denoted by mA .

Definition 2.6.3. Let $(X, \mathcal{A}, \mu)$ be a measure space. A set $B \in \wp X$ is called μ*-negligible* if $\exists A \in \mathcal{A}$ such that $B \subseteq A$ and $\mu(A) = 0$. A property is said to hold μ*-almost everywhere* (usually abbreviated to μ*-a.e.*) if it holds except on a μ-negligible set .

Theorem 2.6.1. Let $(X, \mathcal{A}, \mu)$ be a measure space. If two sets $A \in \mathcal{A}$ and $B \in \mathcal{A}$ are equal μ-a.e., then $\mu(A) = \mu(B)$; moreover, $\forall C \in \mathcal{A}$, $\mu(A \cap C) = \mu(B \cap C)$.

Theorem 2.6.2. Let μ_1 and μ_2 be two measures on a measurable space $(X, \mathcal{A})$. If $\alpha \in R^+$ and $\beta \in R^+$, then $\alpha\mu_1 + \beta\mu_2$ is a measure on the same measurable space .

Definition 2.6.4. Let $(X, \mathcal{A}, \mu)$ be a measure space and $(Y, \mathcal{B})$ be a measurable space. The $\mathcal{A}$-$\mathcal{B}$-measurable function $f : X \to Y$ naturally defines a (positive) measure ν on a $(Y, \mathcal{B})$, such that $\forall B \in \mathcal{B}$:

$$\nu(B) = \mu(f^{-1}B) \quad .$$

Note 2.6.1. Let $(X, \mathcal{A})$ be a measurable space and $(Y, \mathcal{B}, \nu)$ be a measure space. Given a $\mathcal{A}$-$\mathcal{B}$-measurable function $f : X \to Y$, it is not possible to naturally define a (positive) measure μ on a $(X, \mathcal{A})$ **.**

Definition 2.6.5. Let $J \neq \emptyset$ denote an index set, and let $\{(X_j, \mathcal{A}_j, \mu_j)\}_{j \in J}$ be a family of measure spaces. Then, $\forall K \subseteq J$, $K \neq \emptyset$, the *product measure* $\mu_K = \bigotimes_{j \in K} \mu_j$ on the measurable space $(X_K, \mathcal{A}_K)$ is defined, $\forall A_K \in \mathcal{A}_K$ by:

$$\mu_K(A_K) = \prod_{j \in K} \mu_j(p_j^{-1} A_K) \quad .$$

Definition 2.6.6. Let $J \neq \emptyset$ denote an index set, and let $\{(X_j, \mathcal{A}_j, \mu_j)\}_{j \in J}$ be a family of measure spaces. Then, $\forall K \subseteq J$, $K \neq \emptyset$, the triade $(X_K, \mathcal{A}_K, \mu_K)$ is called the *product of the measure spaces* $\{(X_j, \mathcal{A}_j, \mu_j)\}_{j \in J}$, and indicated by:

$$\bigotimes_{j \in K} (X_j, \mathcal{A}_j, \mu_j) \quad .$$

2.7. PROBABILITY MEASURE

Definition 2.7.1. A measure space $(\Omega, \mathcal{E}, P)$ is called a *probability space* if $P(\Omega) = 1$. The set Ω is called *space of events*, the measure P is called *probability measure* or simply *probability*, and a subset $A \subseteq \mathcal{E}$ is called *event* **.**

Example 2.7.1. Let $(X, \wp X)$ be a measurable space, and $a \in X$. If $A \in \wp X$, the *Dirac probability in a of the set A*, $\delta_a A$, is defined by:

$$\delta_a A = \begin{cases} 1 & , \quad \text{if } a \in A \\ 0 & , \quad \text{if } a \notin A \end{cases} \quad .$$

Theorem 2.7.1. Let $(\Omega, \mathcal{E}, P)$ be a probability space. Then, $\forall A \in \mathcal{E}$:

$$P(\Omega - A) = 1 - P(A) \quad .$$

Definition 2.7.2. Let $(\Omega, \mathcal{E}, P)$ be a probability space. An event $A \subseteq \mathcal{E}$ is called *impossible* if $P(A) = 0$; *certain* if $P(A) = 1$ **.**

Theorem 2.7.2. The complementary of an impossible event is certain, and viceversa **.**

Theorem 2.7.3. Let $(\Omega, \mathcal{E}, P)$ be a probability space. If $A \in \mathcal{E}$ and $B \in \mathcal{E}$, such that $\Omega - A \subseteq B$, then:

$$P(A \cap B) \leq P(A) \cdot P(B)$$

Demonstration. Since $\Omega - A \subseteq B$, then $P(A) = P(A \cap B) + P(\Omega - B)$. Now, if $P(B) = 1$, i.e. $P(\Omega - B) = 0$, the thesis is obvious; otherwise, if $P(B) \neq 1$, then:

$$\begin{aligned}
&P(A \cap B) \leq P(B) \\
&P(A \cap B) \cdot P(\Omega - B) \leq P(\Omega - B) \cdot P(B) \\
&P(A \cap B) \cdot [1 - P(B)] \leq P(\Omega - B) \cdot P(B) \\
&P(A \cap B) \leq P(A \cap B) \cdot P(B) + P(\Omega - B) \cdot P(B)
\end{aligned}$$

and, from the first observation:

$$P(A \cap B) \leq P(A) \cdot P(B) \quad .$$

Theorem 2.7.4. Let P_1 and P_2 be two probability measures on a measurable space $(\Omega, \mathcal{E})$. If $\alpha \in [0, 1]$ and $\beta \in [0, 1]$, with $\alpha + \beta = 1$, then $\alpha P_1 + \beta P_2$ is a probability measure on the same measurable space **.**

Definition 2.7.3. Let $(\Omega, \mathcal{E}, P)$ be a probability space. The function $O : \mathcal{E} \to R^+ \cup \{+\infty\}$ defined $\forall A \in \mathcal{E}$ by:

$$O(A) = \begin{cases} \frac{P(A)}{P(\Omega - A)} & , \quad \text{if } 0 \leq P(A) < 1 \\ +\infty & , \quad \text{if } P(A) = 1 \end{cases}$$

is called *odds in favour of the event A* **.**

Definition 2.7.4. Let $(\Omega, \mathcal{E}, P)$ be a probability space and $B \in \mathcal{E}$ an event such that $P(B) \neq 0$. The *B-conditional probability* is the probability measure defined, $\forall A \in \mathcal{E}$, by:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad .$$

Definition 2.7.5. Let $(\Omega, \mathcal{E}, P)$ be a probability space and $B \in \mathcal{E}$ an event such that $P(B) \neq 0$. An event $A \in \mathcal{E}$ is called *B-independent* if $P(A|B) = P(A)$ **.**

Theorem 2.7.5. *Bayes' theorem.* Let $(\Omega, \mathcal{E}, P)$ be a probability space and $\{A_j\}_{j \in J} \subseteq \mathcal{E}$ such that $\bigcup_{j \in J} A_j = \Omega$ and, $\forall j \in J$, $P(A_j) \neq 0$. Then, $\forall B \in \mathcal{E}$ such that $P(B) \neq 0$, $\forall j \in J$:

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i \in J} P(B|A_i)P(A_i)} \quad .$$

Theorem 2.7.6. Let $(\Omega, \mathcal{E}, P)$ be a probability space and $B \in \mathcal{E}$ an event such that $0 \leq P(B) < 1$. Then, $\forall A \in \mathcal{E}$, $P(A|\Omega - B) = 0$ if and only if $P(A) = P(A \cap B)$.

Definition 2.7.6. Let $(\Omega, \mathcal{E}, P)$ be a probability space and $B \in \mathcal{E}$ an event such that $0 < P(B) < 1$. The *likelihood ratio* Lr *of the* B*-conditional probability of the event* A *and the* $\Omega - B$*-conditional probability of the event* A is defined, $\forall A \in \mathcal{E}$ such that $0 < P(A) < 1$, by:

$$Lr(A;B) = \begin{cases} +\infty & , \quad \text{if } P(A|\Omega - B) = 0 \\ \frac{P(A|B)}{P(A|\Omega - B)} & , \quad \text{if } 0 < P(A|\Omega - B) \leq 1 \end{cases} .$$

Theorem 2.7.7. Let $(\Omega, \mathcal{E}, P)$ be a probability space and $B \in \mathcal{E}$ an event such that $0 < P(B) < 1$. Then, $\forall A \in \mathcal{E}$ such that $0 < P(A) < 1$:

$$P(A|B) = \begin{cases} \frac{P(A)}{P(B)} & , \quad \text{if } P(A|\Omega - B) = 0 \\ P(A) \cdot \frac{Lr(A;B)}{1 + P(B)[Lr(A;B) - 1]} & , \quad \text{if } 0 < P(A|\Omega - B) \leq 1 \end{cases} .$$

Theorem 2.7.8. Let $(\Omega, \mathcal{E}, P)$ be a probability space and $A \in \mathcal{E}$ an event such that $0 < P(A) < 1$. Then, $\forall B \in \mathcal{E}$ such that $0 < P(B) < 1$, denoting with $O(B|A)$ the odds in favour of the event B calculated with respect the A-conditional probability:

$$O(B|A) = \begin{cases} +\infty & , \quad \text{if } P(B|A) = 1 \\ Lr(A;B) \cdot O(B) & , \quad \text{if } 0 \leq P(B|A) < 1 \end{cases} .$$

Demonstration. Since the condition $P(B|A) = 1$ implies $P(A) = P(A \cap B)$, the realtion derives from the definitions of odds and likelihood ratio .

Definition 2.7.7. Let $(\Omega, \mathcal{E}, P)$ be a probability space, $(Y, \mathcal{B})$ a measurable space. The function $f : \Omega \to Y$ is called *random variable* if it is $\mathcal{E}$-$\mathcal{B}$-measurable .

Definition 2.7.8. Let $<$ be the irreflexive relation on a partial ordered set $(X, \leq)$, and let $(Y, \mathcal{B})$ be the Borel measurable space generated by the order topology. Let $f : \Omega \to Y$ be a random variable on a probability space $(\Omega, \mathcal{E}, P)$ into the measurable space $(Y, \mathcal{B})$. Then, the *(cumulative) distribution function* P_f *of the random variable* f is defined, $\forall y \in Y$, by:

$$P_f y = P\big(\{\omega \in \Omega | f(\omega) < y\}\big)$$

i.e. $P_f y = P[f^{-1} B_y^-]$.

Note 2.7.1. The probability is defined on the σ-algebra of the events; the distribution function directly on the space of events. .

Definition 2.7.9. A random variable defined on the real Borel measurable space (R, B) is called *real random variable* .

Definition 2.7.10. Let $J \neq \emptyset$ denote an index set, and let $\{(\Omega_j, \mathcal{E}_j, P_j)\}_{j\in J}$ be a family of probability spaces. Then, $\forall K \subseteq J$, $K \neq \emptyset$, the *product probability* $P_K = \bigotimes_{j\in K} P_j$ on the measurable space $(\Omega_K, \mathcal{E}_K)$ is defined, $\forall A_K \in \mathcal{E}_K$ by:

$$P_K(A_K) = \prod_{j\in K} P_j(p_j^{-1} A_K) \quad .$$

Definition 2.7.11. Let $J \neq \emptyset$ denote an index set, and let $\{(\Omega_j, \mathcal{E}_j, P_j)\}_{j\in J}$ be a family of probability spaces. Then, $\forall K \subseteq J$, $K \neq \emptyset$, the triade $(\Omega_K, \mathcal{E}_K, P_K)$ is called the *product of the probability spaces* $\{(\Omega_j, \mathcal{E}_j, P_j)\}_{j\in J}$, and indicated by:

$$\bigotimes_{j\in K} (X_j, \mathcal{E}_j, P_j) \quad .$$

2.8. INTEGRALS

Definition 2.8.1. Let $(X, \mathcal{A}, \mu)$ be a measure space, and let $J \neq \emptyset$ denote a finite index set. If $\phi = \sum_{j\in J} \alpha_j \Phi_{A_j}$ is a measurable positive simple function on X into R^+, then the *integral of* ϕ *with respect to the measure* μ is defined by:

$$\int \phi d\mu = \sum_{j\in J} \alpha_j \mu(A_j) \quad .$$

Definition 2.8.2. Let $(X, \mathcal{A}, \mu)$ be a measure space and $f : X \to R^+$ a measurable function. The *integral of* f *with respect to the measure* μ is defined by:

$$\int f d\mu = \sup_{\phi \leq f} \int \phi d\mu$$

where the functions ϕ are positive simple functions .

Theorem 2.8.1. Let $(X, \mathcal{A}, \mu)$ be a measure space and $f : X \to R^+$ a measurable function. Then, $\forall \lambda \in R^+$:

$$\mu\Big(\{x \in X | f(x) \geq \lambda\}\Big) \leq \frac{1}{\lambda} \cdot \int f d\mu \quad .$$

Definition 2.8.3. Let $(X, \mathcal{A}, \mu)$ be a measure space and $f : X \to R$ a measurable function, such that also its positive (f^+) and negative (f^-) parts are measurable. The *integral of f with respect to the measure* μ is defined by:

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu \quad .$$

Definition 2.8.4. Let $(X, \mathcal{A}, \mu)$ be a measure space and $f : X \to R$ a real measurable function. The function is called *summable* if both the integrals

$$\int f^+ d\mu \quad \text{and} \quad \int f^- d\mu$$

are finite; *integrable* if only one of the two integrals is finite .

Definition 2.8.5. Let $(X, \mathcal{A}, \mu)$ be a measure space and $f : X \to R$ a measurable function. If $A \subseteq X$, the *integral of f over A* is defined by:

$$\int_A f d\mu = \int f \cdot \Phi_A d\mu \quad .$$

Definition 2.8.6. Let f be a real random variable on the probability space $(\Omega, \mathcal{E}, P)$. Then $\forall k \in N$, the *momentum of k-th order of the random variable f* is defined by:

$$E[f^k] = \int f^k d\mu$$

In particular:
(a). for $k = 1$, the quantity $E[f]$ is called *(mathematical) expectation* or *mean value* ;
(b). the quantity $E[(f - E[f])^k]$ is called the *centered momentum of k-th order of the random variable f* ;
(c). the quantity $\sigma^2 = E[(f - E[f])^2]$ is called *variance*, and its square root σ *standard deviation* ;
(d). the function $Z_j = E[e^{f \cdot j}]$ is called *characteristic function* .

Theorem 2.8.2. *Chebyshev inequality.* Let f be a real random variable on the probability space $(\Omega, \mathcal{E}, P)$. Then, $\forall \lambda \in R^+$:

$$P\Big(\{\omega \in \Omega \big| |f - E[f]| \geq \lambda\}\Big) \leq \frac{1}{\lambda^2} \cdot E\Big[\big(f - E[f]\big)^2\Big] \quad .$$

Definition 2.8.7. Let $f : \Omega \to Y$ be a random variable on a probability space $(\Omega, \mathcal{E}, P)$ onto a partial ordered set $(Y, \leq)$. The distribution function P_f is called *regular* if $\exists p_f : Y \to R^+$ such that:

$$P_f y = \int_{u<y} p_f(u)du$$

The function p_f is called *probability density of the random variable* f **.**

Example 2.8.1. Let $n \in N$ be a positive natural number, and $X_n \subseteq \{0\} \cup N$ be the set of all natural numbers from 0 to n, i.e. $X_n = \{0, 1, \ldots, n\}$. On the measurable space $(X_n, \wp X_n)$, $\forall p \in (0, 1)$, the *binomial probability density* is defined, $\forall x \in X_n$, by:

$$p_{n,p}x = \binom{n}{x} p^x (1-p)^{n-x} \quad .$$

Example 2.8.2. Let $X = \{0\} \cup N$ be the set of all non-negative natural numbers. On the measurable space $(X, \wp X)$, $\forall p \in (0, 1)$, the *geometric probability density* is defined, $\forall x \in X$, by:

$$p_p x = p(1-p)^x \quad .$$

Example 2.8.3. Let $X = \{0\} \cup N$ be the set of all non-negative natural numbers. On the measurable space $(X, \wp X)$, $\forall \lambda \in R^+$, the *Poisson probability density* is defined, $\forall x \in X$, by:

$$p_\lambda x = e^{-\lambda} \cdot \frac{\lambda^x}{x!} \quad .$$

Example 2.8.4. Let $X = \{0\} \cup N$ be the set of all non-negative natural numbers. On the measurable space $(X, \wp X)$, $\forall \alpha \in R^+$, the *negative binomial (Pascal) probability density* is defined, $\forall x \in X$, by:

$$p_\alpha x = \binom{\alpha + x - 1}{x} \cdot p^\alpha (1-p)^x \quad .$$

Example 2.8.5. On the real Borel measurable space (R, B), $\forall m \in R$ and $\forall \sigma \in R^+$, the *normal (Gaussian) probability density centered in* m is defined, $\forall x \in R$, by:

$$p_{m,\sigma} x = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] \quad .$$

Example 2.8.6. On the real Borel measurable space (R, B), $\forall s \in R^+$, the *Lorentzian probability density centered in* m is defined, $\forall x \in R$, by:

$$p_s x = \frac{1}{s\pi} \cdot \frac{1}{1 + \frac{x^2}{s^2}} \quad .$$

Example 2.8.7. On the real Borel measurable space (R^+, B), $\forall\lambda \in R^+$, the *exponential probability density* is defined, $\forall x \in R^+$, by:

$$p_s x = \lambda \cdot e^{-\lambda x} \quad .$$

Example 2.8.8. On the real Borel measurable space (R^+, B), $\forall\lambda \in R^+$ and $\forall\alpha \in R^+$,, the *Gamma probability density* is defined, $\forall x \in R^+$, by:

$$p_s x = \frac{\lambda}{\Gamma(\alpha)} \cdot (\lambda x)^{\alpha-1} e^{-\lambda x} \quad .$$

Example 2.8.9. Let $a \in R$ and $b \in R$, $a < b$. On the real Borel measurable space $((a, b), B)$, the *uniform probability density* is defined, $\forall x \in (a, b)$, by:

$$p_{a,b} x = \frac{1}{b-a} \quad .$$

Example 2.8.10. On the real Borel measurable space $((0, 1), B)$, $\forall p \in R^+$ and $\forall q \in R^+$, the *Beta probability density* is defined, $\forall x \in (0, 1)$, by:

$$p_{p,q} x = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \cdot x^{p-1}(1-x)^{q-1} \quad .$$

Example 2.8.11. Let $n \in N$ be a positive natural number. For all $r \leq n$ and $m \leq n$ positive natural numbers, let $X \subseteq \{0\} \cup N$ be the set of all natural numbers from 0 to $\min(r, m)$, i.e. $X = \{0, 1, \ldots, \min(r, m)\}$. On the measurable space $(X, \wp X)$, the *hypergeometric probability density* is defined, $\forall x \in X$, by:

$$p_{n,m,r} x = \frac{\binom{m}{x}\binom{n-m}{r-x}}{\binom{n}{r}} \quad .$$

Definition 2.8.8. Let $J \neq \emptyset$ denote an index set, and let $\{(\Omega_j, \mathcal{E}_j, P_j)\}_{j\in J}$ be a family of probability spaces. Let $\{f_j\}_{j\in J}$ be a family of random variables onto a family of partial order set $\{(Y_j, \leq_j)\}_{j\in J}$, for which $\exists p_{f_J} : Y_J \to R^+$ (so called *joint probability density of the random variable* f_J) such that:

$$P_{f_J} y_J = \int_{u_J <_J y_J} p_{f_J}(u_J) du_J$$

Then, $\forall K \subseteq J$, the *K-marginal of joint probability density* p_{f_J}, p_{f_K} is defined by:

$$p_{f_K}(u_K) = \int_{Y_{J-K}} p_{f_J}(u_K, u_{J-K}) du_{J-K} \quad .$$

REFERENCES

[1]. Karlin Samuel, Taylor Howard M., *A first course in stochastic processes*, Academic Press (1975)

[2]. Rohatgi Vijay K., *An introduction to probability theory and mathematical statistics*, John Wiley & Sons (1976)

[3]. Bauer Heinz, *Probability theory and elements of measure theory*, Academic Press (1981)

[4]. Boccara Nino, *Functional analysis: an introduction for physicists*, Academic Press (1990)

3

IDENTIFICATION AND INDIVIDUALIZATION PROCESSES

In this chapter, the fundamental principles of forensic science are considered, and in particular the concepts of identity, and the processes of identification and individualization.

At first, definitions of the two principal sets (of evidences and of reference population) need; then a causality function which connects them is defined.

Moreover, also the scientific method normally used in forensic science (based on Analysis, Comparison and Evaluation) is analyzed.

3.1. THE SET OF EVIDENCES

The first of the two important sets in forensic science is the set of evidences.

Definition 3.1.1. Let $X \neq \emptyset$ denote the *set of evidences* .

Normally, in forensic science the elements of the set of evidences are bullets, gunshot residues, cartridge cases, ink, soils, drugs, audiotapes, videotapes, glasses, biological material, fingerprints, and generally everything connected to the scene of crime.

Example 3.1.1. Let $X \subseteq N$ be the set of the positive numbers from 1 to 12, i.e. $X = \{1, 2, \ldots, 12\}$. For example, X represents the set of all the possible sums of the outcomes of two dice. A subset $A \subseteq X$, $A = \{4, 11\}$, represents a possible result in the set of evidences .

3.2. THE SET OF REFERENCE POPULATION

The second of the two important sets in forensic science is the set of reference population.

Definition 3.2.1. Let $\Omega \neq \emptyset$ denote the *set of reference population* .

Normally, the elements of the set of reference population are firearms, barrels, people, etc.

Definition 3.2.2. Let $\Omega \neq \emptyset$ be a set of reference population. The *prosecutor's hypothesis* is a subset $C \subseteq \Omega$; its complementary subset $\Omega - C$ is called *defender's hypothesis* .

The subset C represents the set of all the individuals on prosecutor's attention, i.e. the set of the suspect individuals. However, since the choice of the prosecutor's hypothesis is arbitrary, C is normally chosen in order to be the most selective as possible. Actually, if the set C denotes for example a group of firearms, a new more selective prosecutor's hypothesis $\bar{C} \subseteq C$ can be taken in consideration.

Example 3.2.1. Let Ω be the set of all the 36 possible outcomes of two dice. The subset $C \subseteq \Omega$, $C = \{(2,2),(3,5)\}$, represents a prosecutor's hypothesis. A more selective choice is $\bar{C} = \{(2,2)\}$.

3.3. THE CAUSALITY FUNCTION

The main principle of forensic science is that there is a connection between evidences and individuals of the reference population. This link is here represented by a function which connects an element of the set of individual to all the evidences it can produced.

Definition 3.3.1. Let Ω be the set of reference population, and X the set of evidences. The function $f : \Omega \to \wp X$ such that, $\forall \omega \in \Omega$, $f\omega$ represents the subset of X of all the evidences caused by ω, is called *causality function* .

Example 3.3.1. Let Ω be the set of all produced stamps, and let X be the set of all prints on a paper. A causality function can be defined associating to each stamp the set of all the prints on the paper produced by that stamp .

Example 3.3.2. Let Ω be the set of all produced firearms, and let X be the set of all cartridge cases. A causality function can be defined associating to each firearm the set of all the cartridge cases fired by that firearm .

Example 3.3.3. Let Ω be the set of all human biological individuals, and let X be the set of all DNA fragments. A causality function can be defined associating to each human biological individual the set of all the DNA fragments which are from that human biological individual .

Note 3.3.1. As the following two examples show, on the same couple of set Ω and $\wp X$ different causality functions can be defined .

Example 3.3.4. Let Ω be the set of all produced barrels, and let X be the set of all bullets. A causality function can be defined associating to each barrel the set of all the bullets fired by that barrel .

Example 3.3.5. Let Ω be the set of all produced barrels, and let X be the set of all bullets. A causality function can be defined associating to each barrel the set of all the bullets *of the same material* of that of that barrel .

Note 3.3.2. Let $C \subseteq \Omega$ be the prosecutor's hypothesis. The set fC represents the subsets of $\wp X$ of all the evidences associated to all the suspect individuals in the prosecutor's hypothesis .

Note 3.3.3. No previous properties of f (such as injective or surjective) are assumed .

Further typical examples of reference populations, sets of evidences and causality functions are reported in the following table.

Table 3.3.1. *Examples of typical reference populations, sets of evidences and causality functions.*

individuals Ω	**evidences** X	**causality function** f
persons	photos	picturing
	DNA fragments	origin
	blood stains	origin
	voice	origin
	fragments of friction ridge prints	origin
footwears	sole pattern prints	origin
firearms	cartridge cases	firing
firearm barrels	fragments of bullets	firing
inks	prints	production
stamps	prints	production
primer shootings	metallic micro-particles	production
windows	glass fragments	origin
drug standards	chemical substances	classification

Definition 3.3.2. Let $f : \Omega \to \wp X$ be a causality function on a set of reference population Ω, and let $C \subseteq \Omega$ denote a prosecutor's hypothesis. The subset:

$$A_{C,f} = \{x \in X | \exists \omega \in C : x \in f\omega\}$$

is called the *set of all the evidences of C with respect to the causality function f* .

Definition 3.3.3. Let $C \subseteq \Omega$ denote a prosecutor's hypothesis in the set of reference population Ω, and let $A_{C,f}$ be the set of all the evidences of C with respect to the causality function f. A non-empty subset $B \subseteq A_{C,f}$ is called a *test set of C with respect to the causality function f* .

Definition 3.3.4. Let $A \subseteq X$ be a subset in the set of evidences X, and let $f : \Omega \to \wp X$ be a causality function on a set of reference population Ω. The subset $E_{A,f} \subseteq \Omega$ defined by:

$$E_{A,f} = \bigcup_{x \in A} \{\omega \in \Omega | x \in f\omega\}$$

is called the *evidence deduced by A with respect to the causality function f*. For sake of simplicity, let $E_{x,f}$ indicate $E_{\{x\},f}$.

Note 3.3.4. The subset $E_{A,f}$ of the set of reference population Ω is not defined by $f^{-1}A$, but by $E_{A,f} = \bigcup_{x \in A} E_{x,f}$.

Note 3.3.5. The subset $E_{A,f}$ of the set of reference population could be empty .

Example 3.3.6. Let Ω be the set of all outcomes of two dice and $X \subseteq N$ represent the positive numbers from 1 to 12, i.e. $X = \{1, 2, \ldots, 12\}$. Let $f : \Omega \to \wp X$ the causality function which associates to each outcome $(\omega_1, \omega_2) \in \Omega$ the sum of the faces, i.e. $f(\omega_1, \omega_2) = \omega_1 + \omega_2$.
If the prosecutor's hypothesis is $C = \{(2,2),(3,5)\}$, then $fC = \big\{\{4\},\{8\}\big\}$, while $A_{C,f} = \{4, 8\}$; choosing a more selective hypothesis $\bar{C} = \{(2,2)\}$, then $f\bar{C} = \big\{\{4\}\big\}$, while $A_{\bar{C},f} = \{4\}$.
If $A = \{4, 11\}$ is a subset of X, then $E_{A,f} = \{(1,3),(2,2),(3,1),(5,6),(6,5)\}$ represents the evidence; note that $f^{-1}A = \emptyset$.

3.4. IDENTITY

In set theory, the ontological concept of identity (from the Latin *idem*, i.e. "the same") of two elements is based on the possibility of their distinguishing.

Definition 3.4.1. Let $x_1 \in X$ and $x_2 \in X$ be two elements of a set X. The two elements are called *identical* if $x_1 = x_2$.

In this sense, each element is only self-identical, i.e. it is identical only to itself. This concept is also known as *principium identitatis.*
Identity of two elements is also defined epistemologically, i.e. by means of their properties. At first, the concept of "relative identity" is introduced.

Definition 3.4.2. Let D be an equivalence relation on set X. Two elements $x_1 \in X$ and $x_2 \in X$ are called *D-relative identical* if $x_1 D x_2$.

So, an epistemological definition of "absolute identity" can be deduced by hyperbolizing the concept of relative identity.

Definition 3.4.3. Let $\mathcal{D}$ be the set of all equivalence relations on set X. Two elements $x_1 \in X$ and $x_2 \in X$ are called *absolutely identical* if, $\forall D \in \mathcal{D}$, $x_1 D x_2$.

Theorem 3.4.1. Two elements $x_1 \in X$ and $x_2 \in X$ of a set X are identical if only if they are absolutely identical.
Demonstration. Since the banal equivalence relation Δ is the smallest equivalence relation in $\mathcal{D}$, the two elements x_1 and x_2 are absolutely identical if $x_1 \Delta x_2$, i.e., from the definition of the banal equivalence relation, if $x_1 = x_2$. On the contrary, if x_1 and x_2 are identical, then $\forall D \in \mathcal{D}$, since reflexivity is a necessary property of equivalence relations, $x_1 D x_2$, and so x_1 and x_2 are absolutely identical .

In forensic science, this concept is known as the *principle of uniqueness*, according to which "all objects in the universe are unique". In other words, it is always possible to find differences in two diverse objects.
This concept applies either to the set of evidences, and the set of individuals. In the first case, two different bullets, two different biological materials, two different recorded voices are obviously non-identical, and so it is always possible to find differences in them. In the second case, two different firearms, two different finger patterns, two different persons are obviously non-identical, and analogously it is always possible to find differences in them.

3.5. IDENTIFICATION PROCESS

In forensic science, identification process is the first part of the method normally used in forensic laboratories, according to which the characteristics of two different elements of a set can be analyzed and compared. These two steps represent the first part of the standard method, and they are indicated by their acronym "AC".

Definition 3.5.1. Let $\mathcal{D}$ denote the set of all equivalence relations on a set of evidences X. Given $x \in X$, the family of the sets $[x]_D \subseteq X$, $\forall D \in \mathcal{D}$, represents the *family of all the characteristics of x* .

Definition 3.5.2. *Analysis.* Let $\mathcal{D}$ denote the set of all equivalence relations on a set of evidences X. Given $\mathcal{D}_0 \subseteq \mathcal{D}$, the process of establishing, $\forall x \in X$ and $\forall D \in \mathcal{D}_0$, the sets $[x]_D$ is called *analysis* .

Example of typical characteristics in the sets of evidences are reported in the following table.

Table 3.5.1. *Example of characteristics in the sets of evidences.*

X	D
evidences	**characteristic**
photo	pictured biological features height measurements
DNA fragments	DNA profile
blood stains	blood group DNA profile
voice	syllable pronunciation velocity
fragments of friction ridge prints	*minutiæ* position general pattern
cartridge cases	caliber ejection mark
fragments of bullets	caliber striation mark
prints	colour morphology shape
metallic micro-particles	morphology chemical composition
glass fragments	refractive index
chemical substances	chemical composition

According to the principle discussed in the section about identity, two diverse objects can always be different, and otherwise they can belong to the same equivalence class for a subset of equivalence relations.

Definition 3.5.3. Let $x_1 \in X$ and $x_2 \in X$ be two elements of a set X, and let $\mathcal{D}$ denote the set of all equivalence relations on X. The set $\mathcal{D}_{x_1,x_2} \subseteq \mathcal{D}$ denotes the *family of all equivalence relations such that* $x_1 D x_2$ *if* $D \in \mathcal{D}_{x_1,x_2}$.

Note 3.5.1. The set $\mathcal{D}_{x_1,x_2}$ is not-empty (e.g. the equivalence relation $D = X^2$ belongs to $\mathcal{D}_{x_1,x_2}$) .

Note 3.5.2. The set $\mathcal{D}_{x_1,x_2}$ denotes the set of all characteristics x_1 and x_2 have in common; otherwise, its complementary set in $\mathcal{D}$ represents all the characteristics of difference .

Definition 3.5.4. *Comparison.* Let $x_1 \in X$ and $x_2 \in X$ be two elements of a set X, and let $\mathcal{D}$ denote the set of all equivalence relations on X. Given $\mathcal{D}_0 \subseteq \mathcal{D}$, the process of establishing $\mathcal{D}_0 \cap \mathcal{D}_{x_1,x_2}$ is called *comparison* .

Note 3.5.3. Comparison is equivalent to establish, $\forall D \in \mathcal{D}_0$, if $D \in \mathcal{D}_{x_1,x_2}$.

Note 3.5.4. Comparison arises also when we have to classify an object or a substance, and in this case the comparison object is the standard (e.g. heroine, gunshot residues, etc.) .

Example 3.5.1. Let X be the set of all prints on a paper. Two prints x_1 and x_2 can have the same colour, but of different shape .

Example 3.5.2. Let X be the set of all cartridge cases. Two cartridge cases x_1 and x_2 can be of the same caliber, but of different make .

Example 3.5.3. Let X be the set of all ejection marks on cartridge cases. Two ejection marks x_1 and x_2 can have the same shape, but different peculiar marks .

Example 3.5.4. Let X be the set of all fingerprint fragments. Two fingerprint fragments x_1 and x_2 can be both classified as "Arch" for their general shape, but present differences in the position of the *minutiæ* .

According to the above definition, both analysis and comparison are objective steps, since they are based on observation of characteristics: for example, it should be no matter of discussion that two shoeprints present the same pattern, or different marks on the sole.

3.6. INDIVIDUALIZATION PROCESS

The individualization process is the most critical point in forensic science. After identification, elements are classified but where they are from is unknown.
Individualization is the act of evaluating, by comparing two elements in the set of evidences, if they have a common origin in Ω. In other words, the goal is to make a connection between evidences in X and individuals in Ω.

Definition 3.6.1. Let $C \subseteq \Omega$ be a prosecutor's hypothesis in the set of reference population Ω, and let $f : \Omega \to \wp X$ be a causality function. An element $x \in X$ is called *deducible from the prosecutor's hypothesis* C if $x \in A_{C,f}$.

Theorem 3.6.1. Let $C \subseteq \Omega$ be a prosecutor's hypothesis in the set of reference population Ω, and let $f : \Omega \to \wp X$ be a causality function. An element $x \in A_{C,f}$ if and only if $E_{x,f} \cap C \neq \emptyset$.
Demonstration. If $x \in A_{C,f}$, then $\exists \omega \in C$ such that $x \in f\omega$, and so, for definition of $E_{x,f}$, $\omega \in E_{x,f}$; therefore, $E_{x,f} \cap C \neq \emptyset$. On the contrary, if $E_{x,f} \cap C \neq \emptyset$, then $\exists \omega \in E_{x,f} \cap C$ such that, for definition of $E_{x,f}$, $x \in f\omega$, and so $x \in A_{C,f}$.

Definition 3.6.2. Let $f : \Omega \to \wp X$ be a causality function on a set of reference population

Ω. Two elements $x_1 \in X$ and $x_2 \in X$ are called *deducible from a same prosecutor's hypothesis* if $\exists C \subseteq \Omega$ such that $x_1 \in A_{C,f}$ and $x_2 \in A_{C,f}$.

Theorem 3.6.2. Let $f : \Omega \rightarrow \wp X$ be a causality function on a set of reference population Ω. Two elements $x_1 \in X$ and $x_2 \in X$ are deducible from a same prosecutor's hypothesis if and only if both $E_{x_1,f} \neq \emptyset$ and $E_{x_2,f} \neq \emptyset$.
Demonstration. If x_1 and x_2 are deducible from a same prosecutor's hypothesis, then, for the previous theorem and the definition, $\exists C \subseteq \Omega$ such that both $E_{x_1,f} \cap C \neq \emptyset$ and $E_{x_2,f} \cap C \neq \emptyset$, and so both $E_{x_1,f} \neq \emptyset$ and $E_{x_2,f} \neq \emptyset$. On the contrary, if both $E_{x_1,f} \neq \emptyset$ and $E_{x_2,f} \neq \emptyset$, then, defining $C = E_{x_1,f} \cup E_{x_2,f}$, for the definition x_1 and x_2 are deducible from the same prosecutor's hypothesis C .

Definition 3.6.3. Let $C \subseteq \Omega$ denote a prosecutor's hypothesis in the set of reference population Ω, and let $A_{C,f}$ be the set of all the evidences of C with respect to the causality function f. Denoting by $\mathcal{D}$ the set of all equivalence relations on a set of evidences X, the set:

$$\mathcal{D}_{C,f} = \{D \in \mathcal{D} | \forall x_1 \in A_{C,f}, \forall x_2 \in A_{C,f} : D \in \mathcal{D}_{x_1,x_2}\}$$

is the set of all equivalence relations for which $A_{C,f}$ is a subset of a unique class of equivalence .

Note 3.6.1. The set $\mathcal{D}_{C,f}$ is not-empty (e.g. the equivalence relation $D = X^2$ belongs to $\mathcal{D}_{C,f}$) .

Note 3.6.2. *Locard's principle.* In forensic science, the assumption that different characteristics of the elements of the set of evidences X can be due to the fact that they are deducible from diverse individuals in Ω, is known as *Locard's principle* .

Theorem 3.6.3. Let $J \neq \emptyset$ denote a finite index set, and let $\{D_j\}_{j \in J}$ be a family of equivalence relations in $\mathcal{D}_{C,f}$. Then, the intersection $\bigcap_{j \in J} D_j$ is an equivalence relation in $\mathcal{D}_{C,f}$.

Definition 3.6.4. The smallest equivalence relation in $\mathcal{D}_{C,f}$ is indicated by $D_{C,f}$.

Theorem 3.6.4. Let $C \subseteq \Omega$ denote a prosecutor's hypothesis in the set of reference population Ω, and let $A_{C,f}$ be the set of all the evidences of C with respect to the causality function f. Denoting by $D_{C,f}$ the smallest equivalence relation in $\mathcal{D}_{C,f}$, $\forall x \in X$:

$$x \in A_{C,f} \quad \text{if} \quad A_{C,f} \subseteq [x]_{D_{C,f}} \quad .$$

Theorem 3.6.5. Let $C \subseteq \Omega$ denote a prosecutor's hypothesis in the set of reference population Ω, and let $A_{C,f}$ be the set of all the evidences of C with respect to the causality function f. With the previous notation of $\mathcal{D}_{C,f}$, given $y \in A_{C,f}$ and $\mathcal{D}_0 \subseteq \mathcal{D}$, if $x \in A_{C,f}$ and $D \in \mathcal{D}_{C,f}$, then $D \in \mathcal{D}_{xy}$.

Theorem 3.6.6. Let $C \subseteq \Omega$ denote a prosecutor's hypothesis in the set of reference population Ω, and let $A_{C,f}$ be the set of all the evidences of C with respect to the causality function f. Denoting by $\mathcal{D}_{C,f}$ the set of all equivalence relations on a set of evidences X such that $A_{C,f} \subseteq [x]_D$, $\forall D \in \mathcal{D}_{C,f}$, given $x \in A_{C,f}$ and $D \in \mathcal{D}_{C,f}$, then, $\forall B \subseteq A_{C,f}$, $B \subseteq [x]_D$ and, $\forall y \in B$, $D \in \mathcal{D}_{x,y}$.

Definition 3.6.5. The given element $y \in A_{C,f}$ is called *known* or *experimental*; the element $x \in A_{C,f}$ is called *unknown* or *question* .

As in the previous sections, after the ontological aspects the epistemological ones are considered. At first, the method of abduction is introduced.
In logical reasoning, three methods of inference are known: deduction, induction and abduction. Deduction uses a rule (r) and a precondiction (p) in order to make a logical conclusion (c); induction uses the precondition (p) and the conclusion (c) in order to learn a rule (r); finally, abduction uses a conclusion (c) and a rule (r) in order to assume the precondition.

Table 3.6.1. *Differences between three different kinds of logical reasoning.*

	Deduction	**Induction**	**Abduction**
Major hypothesis	rule	precondition	rule
Minor hypothesis	precondition	conclusion	conclusion
Thesis	conclusion	rule	precondition

The typical example is taken from the works of the philosopher Charles Sanders Pierce.
(r). rule: *All the beans in the bag are white* ;
(p). precondiction: *This bean is from the bag* ;
(c). conclusion: *This bean is white* .
Obviously, only the deductive method is the logically correct (e.g. "since all the beans in the bag are white and this bean is from the bag, then this bean is white"), while induction and abduction cannot be formally demonstrated. In particular, induction sounds as: "since this bean is from the bag and this bean is white, then all the beans in the bag are white"; and abduction as: "since all the beans in the bag are white and this bean is white, then this bean is from the bag"!
In forensic science, abduction so is normally used to formulate a precondition, in order to establish the truth of the prosecutor's hypothesis. The method is applied to the previous theorem, where the three sentences are the following:
(r). rule: $y \in A_{C,f}$ and $D \in \mathcal{D}_0$;
(p). precondiction: $x \in A_{C,f}$ and $D \in \mathcal{D}_{C,f}$, and ;

(c). conclusion: $D \in \mathcal{D}_{x,y}$.
So the receipt can be summarized in the following way.

Abduction 3.6.1. Let $C \subseteq \Omega$ denote a prosecutor's hypothesis in the set of reference population Ω, and let $A_{C,f}$ be the set of all the evidences of C with respect to the causality function f. With the previous notation of $\mathcal{D}_{C,f}$, given $y \in A_{C,f}$ and $\mathcal{D}_0 \subseteq \mathcal{D}$, if $D \in \mathcal{D}_{xy}$, then $x \in A_{C,f}$ and $D \in \mathcal{D}_{C,f}$.

Note 3.6.3. If $B \subseteq A_{C,f}$, the theorem and the abduction can be applied $\forall y \in B$.

If no explicit prosecutor's hypothesis is known, then the following theorem holds.

Theorem 3.6.7. Let $x_1 \in X$ and $x_2 \in X$ denote two elements in the set of evidences, and let $E_1 = E_{x_1,f}$ and $E_2 = E_{x_2,f}$ denote the evidenced deduced by x_1 and x_2 with respect to the causality function f, respectively. Given $E_1 \neq \emptyset$ and $\mathcal{D}_0 \subseteq \mathcal{D}$, if $x_2 \in A_{E_1,f}$ and $D \in \mathcal{D}_{E_1,f}$, then $D \in \mathcal{D}_{x_1 x_2}$. Analogously, given $E_2 \neq \emptyset$ and $\mathcal{D}_0 \subseteq \mathcal{D}$, if $x_1 \in A_{E_2,f}$ and $D \in \mathcal{D}_{E_1,f}$, then $D \in \mathcal{D}_{x_1 x_2}$.

In this case, considering:
(r). rule: $E_1 \neq \emptyset$ and $D \in \mathcal{D}_0$;
(p). precondiction: $x_2 \in A_{E_1,f}$ and $D \in \mathcal{D}_{E_1,f}$;
(c). conclusion: $D \in \mathcal{D}_{x_1,x_2}$;
and the analogue ones, the following abduction can be considered.

Abduction 3.6.2. Let $x_1 \in X$ and $x_2 \in X$ denote two elements in the set of evidences, and let $E_1 = E_{x_1,f}$ and $E_2 = E_{x_2,f}$ denote the evidenced deduced by x_1 and x_2 with respect to the causality function f, respectively. Given $E_1 \neq \emptyset$ and $\mathcal{D}_0 \subseteq \mathcal{D}$, if $D \in \mathcal{D}_{x_1 x_2}$, then $x_2 \in A_{E_1,f}$ and $D \in \mathcal{D}_{E_1,f}$. Analogously, given $E_2 \neq \emptyset$ and $\mathcal{D}_0 \subseteq \mathcal{D}$, if $D \in \mathcal{D}_{x_1 x_2}$, then $x_1 \in A_{E_2,f}$ and $D \in \mathcal{D}_{E_1,f}$.

Definition 3.6.6. *Evaluation.* In the hypotheses of the previous abduction, the process of establishing if $\mathcal{D}_0 \cap \mathcal{D}_{x_1,x_2}$ implies that if $D \in \mathcal{D}_{xy}$, then $x \in A_{C,f}$ and $D \in \mathcal{D}_{C,f}$, in the first case, or $x_2 \in A_{E_1,f}$ and $D \in \mathcal{D}_{E_1,f}$, or $x_1 \in A_{E_2,f}$ and $D \in \mathcal{D}_{E_1,f}$, in the second case, is called *evaluation* .

Between the equivalence relations, someone is trivial, not useful or useful. Two fingerprints can both be green, and have an "Arch" shape. If the interest is to detect if they are from the same finger, their colour is obviously not important while their characterization as "Arch" is relevant. Otherwise, if the interest is in the kind of ink they are produced by, obviously the colour is important while the "Arch" shape is irrelevant. So, evaluation is the most critical step in the standard method: differences and analogies arisen from the comparison must be evaluated in decision making.
Moreover, since the process of evaluation is not deductive but abducted, it is not objective. In other words, each forensic scientist can subjectively evaluate if the same characteristics

of x_1 and x_2 in the set $\mathcal{D}_0$ are enough to abduct that there exists a prosecutor's hypothesis from which x_1 and x_2 are deducible.
The standard method based on the three steps "analysis", "comparison" and "evaluation" is indicated in literature by his acronym "ACE".

Definition 3.6.7. *False positive.* A positive abduction even if in reality $x \notin A_{C,f}$ in the first case, or $x_2 \notin A_{E_1,f}$, or $x_1 \notin A_{E_2,f}$, in the second case, is called *false positive* .

Definition 3.6.8. *False negative.* A negative abduction even if in reality $x \in A_{C,f}$ in the first case, or $x_2 \in A_{E_1,f}$, or $x_1 \in A_{E_2,f}$, in the second case, is called *false negative* .

Note 3.6.4. The process of artificially creating an elements $x \in X$ so that it is considered a false positive, is called *simulating*; on the contrary, the process of artificially creating an elements $x \in X$ so that it is considered a false negative, is called *dissimulating* .

Even if subjective, evaluation is based on standards, and there are generally accepted procedures and rules a good competent examiner will follow.
Note that the process of individualization can have different values in the proceedings. Actually, normally connection between the crime scene and the evidences, and between the evidences and the suspects are considered.
Sometimes, the connection between the crime and the evidences is very high: it is the cases, for example, of the videotapes where a crime is recorded, or an audiotape where menace is recorded. There, the image or the voice are the evidences, and they are nearly linked to the crime. The connection between the evidences and the suspect becomes so very important in decision making. In this case, the report of the forensic scientist is very fundamental for the evaluation of the posterior odd.
In other cases, the evidence can be nearly connected to the suspect, for example in the case of identification by wide fragments of fingerprint or of DNA. But it is not so obvious that the evidence is linked to the crime: for fingerprint, the suspect could be on the scene of crime before or after the crime, or during the crime without being connected to the crime itself. In this sense, DNA is a weaker evidence since it do not prove the presence on the crime scene of the suspect: DNA could reach the crime scene independently from the suspect.
The forensic science is normally interested in the evaluation of the second step (connection between the evidences and the suspect), while the investigator in the first one (connection between the evidences and the crime). Finally, the judge have to collect the two information together and express his decision.

Example 3.6.1. *Personal individualization.* If the set of individuals Ω is constituted by people, the process of individualization is *personal individualization* .

Different kinds of characteristics can be found in literature.

Definition 3.6.9. *Class characteristics.* Characteristics for which, $\forall x \in X$, the evidence $E_{[x]_D,f}$ contains a great number of individuals, are called *class characteristics* .

Definition 3.6.10. *Individual characteristics.* Characteristics for which, $\forall x \in X$, the evidence $E_{[x]_D,f}$ contains few (or one) individuals, are called *individual characteristics* .

In other words, class characteristic refers to general and limited properties of the individuals; while individual ones to peculiar properties of the individuals.

Example 3.6.2. In the set of evidences X of cartridge cases, the characteristic D indicating the caliber is a class characteristic; a peculiar sign in the ejection mark can be a an individual characteristic .

REFERENCES

[1]. Falco Giuseppe, *Identità. Metodo scientifico di segnalamento e identificazione*, Maglione & Strini (1923)

[2]. Sorrentino Ugo, *La scienza contro il crimine* (1950)

[3]. Tuthill Harold, *Individualization: principles and procedures in criminalistics*, Lightning Powder Company Inc. (1994)

[4]. Manning Alan D., *What is Explanation?*, IEEE Trans. Prof. Commun., 44 (2), 53-57 (2001)

[5]. Pàstena Pietro, *La scienza delle tracce. L'identificazione scientifica dell'autore di un crimine*, Bonanno Editore (2003)

[6]. Rizzi Alfredo, *Abduzione ed inferenza statistica*, to be printed on: Statistica e società (2004)

4

PROBABILITY IN FORENSIC SCIENCE

In the previous chapter, the individualization process has been discussed, according to which it is possible to abductively associate evidences to individuals of a reference population. The opposite has not been discussed: once the evidences are compatible with a prosecutor's hypothesis C, it should be questioned if and how many other hypotheses are compatible with the evidences. In order to answer to this question, different reasoning can be done, but the answer often present in the Courtroom is often based on good sense and experience of the forensic expert. Considering this approach not enough in order to have a satisfactory answer, a statistical approach can be introduced.

In forensic science, the widespread accepted approach in the evaluation of evidences is the probabilistic one, based on the *Bayesian model*. In this approach, the judge have to discriminate two complementary hypotheses in probabilistic version, by estimation of the odds, i.e. the ratio between the their probabilities: as in gambling, the winning probability with respect to the loosing one is the measurement of decision making of what to do. And the forensic science has to support the judge in decision making by evaluating of the likelihood ratio between conditional probabilities.

For sake of simplicity, all subscripts indicating the prosecutor's hypothesis C and the causality function f will be omitted, when no ambiguity can arises (e.g. E will indicate $E_{C,f}$, etc.).

4.1. PROBABILITY SPACE

By introducing a probability measure on the set of reference population Ω, some definitions of the previous chapter can be reviewed.

Definition 4.1.1. Let $\{\Omega, \mathcal{E}, P\}$ be a probability space on the set of reference population Ω .

Definition 4.1.2. Let the prosecutor's hypothesis C be a measurable set on the probability space $\{\Omega, \mathcal{E}, P\}$, i.e. $C \in \mathcal{E}$.

In this scenario, the probability of C, $P(C)$, and the probability of the defender's hypothesis $\Omega - C$, $P(\Omega - C) = 1 - P(C)$ can be calculated. These probabilities are called *prior probabilities*, and, if $P(C) < 1$, their ratio:

$$O(C) = \begin{cases} \frac{P(C)}{P(\Omega - C)} & , \quad \text{if } 0 \leq P(C) < 1 \\ +\infty & , \quad \text{if } P(C) = 1 \end{cases}$$

represents the *prior odds* in favour of the prosecutor's hypothesis.
Normally, the judge is at first interested in evaluation of the prior odds $O(C)$. If $O(C) > 1$, the prior probability of the prosecutor's hypothesis is greater that the prior probability of the defender's one. Normally, an appropriate choice of C can assure $O(C) < 1$.

Note 4.1.1. A causality function f on the probability space $\{\Omega, \mathcal{E}, P\}$ is a random variable, and so on $\wp X$ a probability space $\{\wp X, \mathcal{A}, \Pi\}$ is naturally defined .

As described in the previous chapter, individualization allows the forensic expert to define in the set of reference population Ω the evidence E.

Definition 4.1.3. Let the evidence E be a measurable set on the probability space $\{\Omega, \mathcal{E}, P\}$, i.e. $E \in \mathcal{E}$.

4.2. BAYESIAN APPROACH

When an evidence E, with $P(E) \neq 0$, is considered, this new information in the scenario requires an up-date of the calculation of the probabilities the judge has now to take in account: the so called *posterior probabilities* are calculated by conditioning C and $\Omega - C$, respectively, by the evidence E. Consequently, their *posterior odds* can be estimated:

$$O(C) = \begin{cases} \frac{P(C|E)}{P(\Omega - C|E)} & , \quad \text{if } 0 \leq P(C|E) < 1 \\ +\infty & , \quad \text{if } P(C|E) = 1 \end{cases}$$

Definition 4.2.1. The quantity $P(E|C)$ is called *match probability* .

Definition 4.2.2. The quantity $P(E|\Omega - C)$ is called *random match probability* .

According to Bayes's theorem, the two posterior and the prior odds are proportional by a multiplicative factor $Lr(E;C)$, called *likelihood ratio* or *Bayes factor*, which is defined as the ratio between the match probability and the random match probability: if $0 < P(C) < 1$ and $0 < P(E) < 1$, then:

$$O(C|E) = \begin{cases} +\infty & , \quad \text{if } P(C|E) = 1 \\ Lr(E;C) \cdot O(C) & , \quad \text{if } 0 \le P(C|E) < 1 \end{cases}$$

where:

$$Lr(E;C) = \begin{cases} +\infty & , \quad \text{if } P(E|\Omega - C) = 0 \\ \frac{P(E|C)}{P(E|\Omega - C)} & , \quad \text{if } 0 < P(E|\Omega - C) \le 1 \end{cases}$$

The aim of the forensic scientist lies in the calculation of the likelihood ratio, in order to support the judge in the evaluation of the posterior odds. Accordingly, the forensic scientist has to present the results of his analysis in such appropriate way that no doubt of interpretation could arise.
If $Lr(E;C) > 1$, the evidence E favours the prosecutor's hypothesis, and the greater the value of the likelihood ratio is, the more determinant the evidence is. On the contrary, if $Lr(E;C) < 1$, the evidence E increases the defender's hypothesis probability rather than the prosecutor one, and $Lr(E;C) \ll 1$ means that the evidence makes the defender's hypothesis very strong. Finally, $Lr(E;C) = 1$ implies the absolutely unconcern of the evidence about the two hypotheses.
The result can also be indicated by means of the base-10 logarithm of the likelihood ratio, whose sign indicates if which of the two hypotheses is strong (plus, for the prosecutor's hypothesis, minus for the defender's one), and whose value gives specifies the relative magnitude order of the strength of the hypothesis.

Example 4.2.1. Let E an impossible event, i.e. $P(E) = 0$. Then, the evidence E is *absurd* .

Example 4.2.2. Let E a certain event, i.e. $P(E) = 1$. Then, since $P(E|C) = 1$ and $P(E|\Omega - C) = 1$, $Lr(E;C) = 1$, i.e. the evidence is *non-essential* .

Example 4.2.3. Let E an event such that $E \subseteq C$. Then, since $P(E|C) \neq 0$ and $P(E|\Omega - C) = 0$, $Lr(E;C) = +\infty$, i.e. the evidence is *essential, in favour of C* .

Example 4.2.4. Let E an event such that $E \subseteq \Omega - C$. Then, since $P(E|C) = 0$ and $P(E|\Omega - C) \neq 0$, $Lr(E;C) = 0$, i.e. the evidence is *essential, in favour of* $\Omega - C$.

If finite, the likelihood ratio can also be written:

$$Lr(E;C) = \frac{P(E|C) \cdot [1 - P(C)]}{P(E) - P(E|C) \cdot P(C)}$$

or

$$Lr(E;C) = \frac{P(E \cap C)}{P(E) - P(E \cap C)} \cdot \frac{1 - P(C)}{P(C)}$$

These expressions lighten the three dimensional character of the likelihood ratio: the exact calculation needs the evaluation of three independent parameters: $P(E)$, which represents the probability of the evidences, and which is normally estimated by means of databases; the match probability $P(E|C)$ (or alternatively $P(E \cap C)$); and the prior probability $P(C)$, which is often unknown to the forensic expert.
Moreover, given $P(E)$ and $P(E|C)$, it is not possible in general to deduce $P(E|\Omega - C)$.

Example 4.2.5. In the case of individualization by means of DNA profiling, normally the *loci* of the evidence E can or cannot match with the suspect DNA code. In the first case, $P(E|C) = 1$, while in the second $P(E|C) = 0$. The statistical frequencies of the *loci* in the human population are scheduled ($P(E)$). But, it is not possible to know the exact value of $P(E|\Omega - C)$ $\quad$ **.**

Example 4.2.6. Let Ω be the set of all outcomes of two dice and $X \subseteq N$ represent the positive numbers from 1 to 12, i.e. $X = \{1, 2, \ldots, 12\}$. Let $f : \Omega \to \wp X$ the random variable which associates to each outcome $(\omega_1, \omega_2) \in \Omega$ the sum of the faces, i.e. $f(\omega_1, \omega_2) = \omega_1 + \omega_2$. Let all the events be considered with equal probability on the measurable space $(\Omega, \wp\Omega)$, i.e. the probability measure such that $P(\{\omega\}) = 1/36$, $\forall \omega \in \Omega$. If the prosecutor's hypothesis is $C = \{(2,2), (3,5)\}$, then $P(C) = 1/18$ and $P(\Omega - C) = 17/18$, and the prior odds are $O(C) = 1/17$.
If the set $A \subseteq X$ is $A = \{4, 11\}$, then $E = E_{A,f} = \{(1,3), (2,2), (3,1), (5,6), (6,5)\}$ represents the evidence, and $P(E) = 5/36$.
Moreover, $P(E|C) = 1/2$, $P(E|\Omega - C) = 2/17$ and $Lr(E;C) = 17/4$. Again, $P(C|E) = 1/5$ and $P(\Omega|E) = 4/5$, and the posterior odds are $O(C|E) = 1/4$. But, it is not possible to deduce the last result from the two known values $\quad$ **.**

In the hypotheses $C \subseteq E$ or $\Omega - C \subseteq E$, the inequalities of the following examples hold. Note that often C can be choose so that one of the previous two hypotheses is verified.

Example 4.2.7. Let E an event such that $C \subseteq E$. Then, since $P(E|C) = 1$:

$$Lr(E;C) = \frac{1}{P(E|\Omega - C)} \geq \frac{1}{P(E)} \geq 1 \quad .$$

Example 4.2.8. In the case of individualization by means of DNA profiling, if $P(E|C) = 1$, then $C \subseteq E$, and hence $\frac{1}{P(E)}$ is an underestimation of $Lr(E;C)$ $\quad$ **.**

Example 4.2.9. Let E an event such that $\Omega - C \subseteq E$. Then, since $P(E|\Omega - C) = 1$:

$$Lr(E;C) = P(E|C) \leq P(E) \leq 1 \quad .$$

4.3. ERRORS IN INTERPRETATION

Examples of fallacies in interpreting probabilistic results abound.

Fallacy 4.3.1. *Prosecutor's fallacy.* The *prosecutor's fallacy*, or *transposed conditional fallacy*, consists in considering the value calculated $P(E|\Omega - C)$ instead of the value of $P(\Omega - C|E)$.

Example 4.3.1. In the example of the two dice, the probability of the evidence E in the defender hypothesis is $P(E|\Omega - C) = 2/17 = 11.8\%$. The prosecutor's fallacy considers that the posterior probability $P(\Omega - C|E)$ holds 11.8%, and so the posterior probability of C, $P(C|E)$, is 89.2% .

Fallacy 4.3.2. *Defender's fallacy.* The *defender's fallacy* consists in considering the value of the evidence E irrelevant since the $O(C|E) \ll 1$, even if $Lr(E;C) > 1$.

Example 4.3.2. In the example of the two dice, since $O(C|E) = 1/4 = 25\%$ is considered by the defender not enough in order to proof C, he consider irrelevant the evidence E, even if $Lr(E;C) = 17/4 = 4.25\%$. The fact that $O(C|E) < 1$ depends by the low value of the prior odds: $O(C) = 1/17 = 5.9\%$.

In literature, famous cases in which fallacies was erroneously considered in decision are reported (e.g. the cases of Dreyfus, *People v. Collins*).

4.4. VIRTUAL SCALE OF INTERPRETATION

In some branches of forensic science, since no accepted mathematical models are in used in evaluation of the probabilities in Ω, approximated *virtual scales* of interpretation are used. Different degrees of interpretation can be used, even if the virtual scales are normally based on four different levels, as reported in the following table.

Table 4.4.1. *Example of a four value virtual scale.*

result	likelihood ratio	comments
high positivity	$Lr(E;C) \gg 1$	class and individuals characteristics match
low positivity	$Lr(E;C) > 1$	class characteristics match, no relevant individual characteristics
inconclusive	$Lr(E;C) = 1$	no relevant class characteristics, no relevant individual characteristics
negative	$Lr(E;C) < 1$	class or individual characteristics do not match

The trying of introduction in automatic method in comparison is developing algorithms which allow a quantification of the match probability.

4.5. APPLICATIONS

Many applications of Bayesian approach in forensic science are known in literature.
The first example is in the so called *sampling theory*, generally used when it is not possible or it is not economically moderate to analyze all the evidences. It is the case of an analysis of a great number of tablets of suspect drug, of ammunition, or of glass fragments. Again, sampling theory is used when diverse exclusive-destroying analyses are need on the exhibit, and so the prosecutor intends to make different examinations on group of evidences.
Secondly, applications on personal individualization are very important in forensic science. Actually, personal individualization allows a direct connection between the evidences and people who are involved in legal proceedings; on the contrary, in applications on object individualization, although non trivial in legal proceedings, a direct link between objects and involved people has to be demonstrated.
Applications in personal individualization are based on DNA analysis, on voice recognition, on friction ridge prints (fingerprints, palmprints and footprints), and on somatic features. Different levels of individualization can normally be reached: DNA and friction ridge prints, in the case of evidences with significant information, can imply very high values of the likelihood ratio; on the contrary, with voice recognition and somatic features comparison, lower values of $Lr(E;C)$ are reached. But, as previously noted, sometimes in these cases the evidence is strictly connected to the crime, while friction ridge prints and DNA do not. Actually, friction ridge print demonstrates the presence of the person on the scene of crime, but not his direct connection to the crime; DNA do not demonstrate the presence of the person since, for example, a cigarette boot can be present on the scene of crime by different causes. In this sense, the results of the use of police database in which fingerprints and DNA profiles are filed (AFIS, CODIS, etc.) should always be considered by the judge in accordance to intelligence results. Note that, since DNA follows well-known heredity rules of transmission from parents, it is useful for recognition of missing persons by comparison with DNA profiles of relatives.
Finally, applications on object individualization (glasses, fibers, etc.), and in calculating their transfer probabilities, are known in literature.

REFERENCES

[1]. Aitken Colin G.G., *Statistics and the evaluation of evidence for forensic scientists*, John Wiley & Sons (1997)

[2]. Karlin Samuel, *A first course in stochastic processes*, Academic Press (1975)

[3]. Zavattaro Davide, *L'identificazione e il problema dell'interpretazione del dato scientifico. Riflessi dibattimentali*, Rassegna dell'Arma dei Carabinieri 4, 53-65 (2001)
[4]. Pàstena Pietro, *La scienza delle tracce. L'identificazione scientifica dell'autore di un crimine*, Bonanno Editore (2003)
[5]. Gastwirth Joseph L., *Statistical science in the Courtroom*, Springer-Verlag Telos (2000)
[6]. Good Phillip I., *Applying statistics in the Courtroom*, Chapman & Hall/CRC (2001)
[7]. Azoury Myriam, Grader-Sageev David, Avraham Shlomit, *Evaluation of a sampling procedure for heroin street doses*, J Forensic Sci, 43, 6, 1203-1207 (1998)
[8]. Aitken Colin G.G., *Sampling - How big a sample?*, J Forensic Sci, 44, 4, 750-760 (1999)
[9]. Amraoui Yacine, Allio Ingrid, Garcia Carine, Perrin Martine, *Échantillonage et interprétation: application aux produits de saise analysés par un laboratoire de toxicologie*, Annales de Toxicologie Analytique, XIII, 4, 265-274 (2001)
[10]. Curran James M., Triggs C.M., Buckleton Jhon S., *Sampling in forensic comparison problems*, Science & Justice, 38, 2, 101-107 (1998)
[11]. Tzidony Dov, Ravreby Mark, *A statistical approach to drug sampling: a case study*, J Forensic Sci, 37, 6, 1541-1549 (1992)
[12]. Frank Richard S., Hinkley Sidney W., Hoffman Carolyn G., *Representative sampling of drug seizures in multiple containers*, J Forensic Sci, 36, 2, 350-357 (1991)
[13]. *Guidelines on representative drug sampling*, European Network of Forensic Science Institutes, Drug Working Group (2004)
[14]. Evett Ian W., Weir Bruce S., *Interpreting DNA evidence*, Sinauer Associates (1998)
[15]. Vanezis Peter, Lu David, Cockburn Jeff, Gonzalez America, McCombe Graeme, Trujillo Orlando, Vanezis Maria, *Morphological classification of facial features in adult Caucasian males based on an assessment of photographs of 50 subjects*, J Forensic Sciences, 41 (5), 786-791 (1996)
[16]. Stoney David A., Thorton John I., *A critical analysis of quantitative fingerprint individuality models*, Journal of Forensic Sciences, JFSCA, 31 (4) 1187-1216 (1986)
[17]. Curran James M., Hicks Tacha N., Buckleton Jhon S., *Forensic interpretation of glass evidence*, CRC Press (2000)
[18]. Causin Valerio, Schiavone Sergio, Marigo Antonio, Carresi Pietro, *Bayesian framework for evaluation of fiber evidence in a double murder - A case report*, For Sci Int 141, 159-170 (2004)

5

THE EXTENDED LIKELIHOOD RATIO

The use of Bayesian approach in likelihood evaluation failed when the evidence E and the prosecutor's hypothesis C are disjoint events. This means that the evidence is not compatible with the suspect, and then the casework has a negative conclusion. The reason is trivially in the fact that the value of the standard definition of conditional probability $P(A|B)$ is zero when the two events A and B are disjoint. This non-compatibility of E and C can be very common, especially using continuous variables (e.g. height, voice frequencies, etc.), where the evaluation of the odds fails. Moreover, the difference should be evaluated with respect to a database, in order to have different evaluation in the center of the probability density of the population rather than in the tail.

For example, if the suspect's height measure in centimeters is $C = \{180\}$, $P(C|E) = 0$ if the evidence is both $E = \{179\}$ and $E = \{210\}$; but the first case could be not exclusive as the second one since the difference with C is just 1 centimeter. Moreover, if $C = \{209\}$ and $E = \{210\}$ the rarity of the measurements of C in the population should bring to a more conclusive case rather than the case in which $C = \{180\}$ and $E = \{179\}$.

In this chapter, a new approach to calculate the probability of the intersection of two events is proposed. It is depending on the database and a positive parameter ε, such that the evaluation of the likelihood ratio is not zero when E and C are disjoint, and it reduces to the standard case in the limit $\varepsilon \to 0$.

This result is achieved by construction of a copula and a function $\lambda(\varepsilon)$ is also evaluated in order to control the concentration properties of the copula.

5.1. THE EXTENDED PROBABILITY

Let X denote the random variable which models a characteristic of the population in the probability space $(\{\Omega, \mathcal{E}, P\}$ (e.g. the height measure), and let P_X indicate the relative probability measure.
Since the standard formula for the likelihood ratio:

$$Lr(E;C) = \frac{P_X(E \cap C)}{P_X(E \cap \Omega - C)} \cdot \frac{P_X(\Omega - C)}{P_X(C)}$$

involves the evaluation of the probability of the intersection of two sets, the purpose is to introduce a new approach to the computation of the probability so that the intersection of two disjoint events A and B is not trivially nought.
Let $P^{(\varepsilon)}_{XX}$ define a suitable family of bivariate probability measures on Ω^2, depending on a positive parameter ε, such that, $\forall A \in \mathcal{E}$ and $\forall B \in \mathcal{E}$, the following properties hold:

$$\begin{aligned} P^{(\varepsilon)}_{XX}(A, B) &> 0, \quad \text{if} \quad A, B \neq \emptyset \\ P^{(\varepsilon)}_{XX}(A, B) &= P^{(\varepsilon)}_{XX}(B, A) \\ P^{(\varepsilon)}_{XX}(A, \Omega) &= P_X(A) \\ \lim_{\varepsilon \to 0} P^{(\varepsilon)}_{XX}(A, B) &= P_X(A \cap B) \end{aligned}$$

This means that the bivariate probability is positive, symmetric, its marginals coincide to the univariate probability, and in the limit $\varepsilon \to 0$ it reduces to the the standard definition of the probability of the intersection of the two events.
The proposed method in generalizing the formula of the likelihood ratio is based on the following rule:

to replace in the likelihood ratio the univariate probability of the intersection $P_X(A \cap B)$ *with the extended bivariate probability* $P^{(\varepsilon)}_{XX}(A, B)$.

More precisely, the replacement is:

$$P_X(A \cap B) \quad \to \quad P^{(\varepsilon)}_{XX}(A, B)$$

According to this scheme, the likelihood ratio can be replaced by:

$$Lr^{(\varepsilon)}(E;C) = \frac{P^{(\varepsilon)}_{XX}(E, C)}{P^{(\varepsilon)}_{XX}(E, \Omega - C)} \cdot \frac{P_X(\Omega - C)}{P_X(C)}$$

which tends to the standard one in the limit $\varepsilon \to 0$.
The construction of the extended probability measure $P^{(\varepsilon)}_{XX}$ can be based on the theory of copulas. A suitable copula is constructed by a bivariate Gaussian distribution with ε-depending covariance, and using this copula, the bivariate distribution has as marginals the chosen database, assumed to have a positive density.

5.2. GENERALITIES

Some basic facts on bivariate random vectors are here recalled.
Given a real random variable X and denoted its probability density by p_X, supposed to be sufficiently regular and positive, the probability of the event $A \subseteq R$ is:

$$P_X(A) = \int_A ds\ p_X(s)$$

and the distribution function is:

$$F_X(x) = \int_{s<x} ds\ p_X(s)$$

If (X, Y) is a pair of random variables,let p_X and p_Y denote the two probability densities, respectively, and let p_{XY} be their joint probability density. With these notations, the probability calculated in the rectangle (A, B) of R^2 with bases A and B is:

$$P_{XY}(A, B) = \int_A ds \int_B dt\ p_{XY}(s,t)$$

The relationships between the joint density and its marginals are the following:

$$p_X(s) = \int_\Omega dt\ p_{XY}(s,t)$$
$$p_Y(t) = \int_\Omega ds\ p_{XY}(s,t)$$

5.3. CONSTRUCTION OF A COPULA

In order to define the family $P^{(\varepsilon)}_{XX}$, standard arguments related to the theory of copulas are used.
Essentially, a copula is a bivariate probability distribution $K(u, v)$ on $[0, 1]^2$ such that its marginals are uniform on $[0, 1]$.
The cited problem is equivalent to construct a family of copulas $K^{(\varepsilon)}(u, v)$, depending on the positive parameter ε, which are concentrated along the diagonal $u = v$.

The main tool is the inversion method based on the well-known property: if Z is a random variable and F_Z is its distribution function (supposed to be invertible), then:

$$F_Z(Z) = U$$

where U is the uniform random variable on the interval $[0, 1]$. So, the pair of random variables (U_1, U_2) defined as:

$$\begin{aligned} U_1 =& F_{Z_1}(Z_1) \\ U_2 =& F_{Z_2}(Z_2) \end{aligned}$$

has a joint probability distribution which is a copula.
A possible choice is to use a bivariate Gaussian vector (Z_1, Z_2) with zero mean and covariance matrix C_ε:

$$C_\varepsilon = \begin{bmatrix} 1 & 1-\varepsilon \\ 1-\varepsilon & 1 \end{bmatrix}$$

[note that $\det C_\varepsilon = \varepsilon(2-\varepsilon)$].
If $\mathbf{z} = (z_1, z_2)$, the joint probability density is:

$$\varphi_\varepsilon(\mathbf{z}) = \frac{1}{2\pi \cdot det^{1/2} C_\varepsilon} \cdot \exp\left(-\frac{1}{2}\mathbf{z}C_\varepsilon^{-1}\mathbf{z}\right)$$

where C_ε^{-1} denotes the inverse of the covariance matrix. Let $\Phi^{(\varepsilon)}$ denote the associated distribution function.
With this choice of φ_ε, the function $\varphi_0(z_1, z_2)$ can be defined as:

$$\varphi_0(z_1, z_2) = \lim_{\varepsilon \to 0} \varphi_\varepsilon(z_1, z_2) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}z_1 z_2\right) \cdot \delta(z_1 - z_2)$$

where the limit $\delta(x) = \lim_{\sigma \to 0} \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right)$ has been used. Moreover:

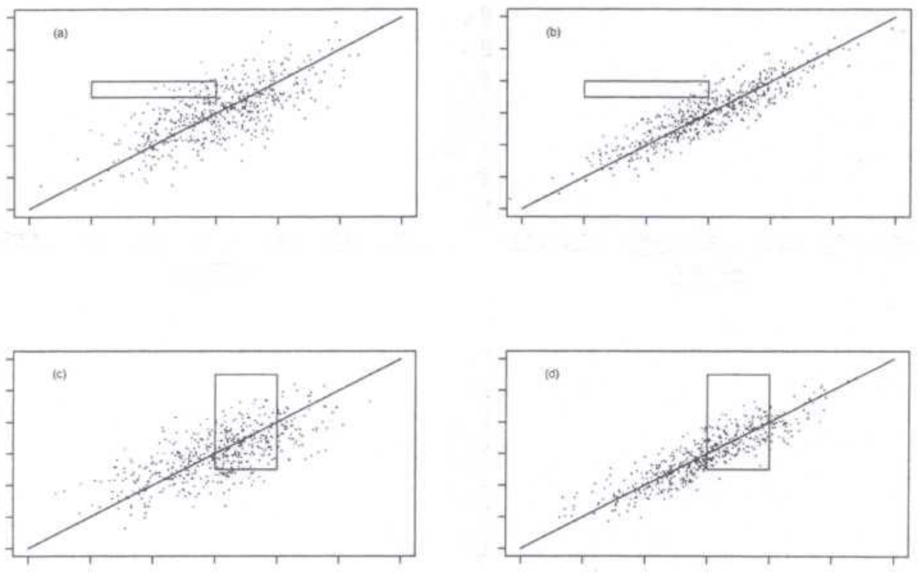
$$\varphi_1(\mathbf{z}) = \frac{1}{2\pi} \cdot \exp\left(-\frac{1}{2}\mathbf{z}^2\right)$$

i.e. the case of Z_1 and Z_2 independent.
Since the two marginal densities have the same ε-independent form, this can be written by:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}z^2\right)$$

and, in the same way, $\Phi(z)$ is the distribution function.
The function $\varphi_\varepsilon(z_1, z_2)$ represents a bivariate density with the mass concentrated on the diagonal line $\{z_1 = z_2\}$ in R^2, and centered in $(0, 0)$, as shown in the following figure. The rectangle (A, B) has vanishing contribute to the probability as $\varepsilon \to 0$, as expected.

Figure 5.2.1. *Simulation of a bivariate density with 500 points, using the free statistical software* R. *(a) and (b): rectangle not intersecting the diagonal for* $\varepsilon = 0.3$ *and* $\varepsilon = 0.1$*, respectively; (c) and (d): rectangle intersecting the diagonal for* $\varepsilon = 0.3$ *and* $\varepsilon = 0.1$*, respectively.*



A copula can be defined by:

$$K^{(\varepsilon)}(u,v) = \Phi^{(\varepsilon)}(\Phi^{-1}(u), \Phi^{-1}(v))$$

It is well-known that if X and Y are two real random variables with distribution F and G respectively, and K is a copula, then the function H defined by:

$$H(x,y) = K(F(x), G(y))$$

is a bivariate distribution function with marginals F and G.
Applying this property to the pair (X, X), where X is the database, the family of $\Phi^{(\varepsilon)}$ defines a family $H^{(\varepsilon)}_{XX}$:

$$H^{(\varepsilon)}_{XX}(x,y) = K^{(\varepsilon)}(F_X(x), F_X(y))$$

such that each of them is a bivariate distribution function with both marginals F_X. The densities $h^{(\varepsilon)}_{XX}$ can be computed according to the formula:

$$h^{(\varepsilon)}_{XX}(x,y) = \frac{\partial^2 H^{(\varepsilon)}_{XX}(x,y)}{\partial x \partial y}$$

Finally, the probability measure $P^{(\varepsilon)}_{XX}(A,B)$ can be defined by:

$$P^{(\varepsilon)}_{XX}(A,B) = \int_A ds \int_B dt \; h^{(\varepsilon)}_{XX}(s,t)$$

Making a suitable change of variables:

$$u = F_X(s) \quad \text{and} \quad v = F_X(t)$$

the probability $P^{(\varepsilon)}_{XX}(A,B)$ can be written as:

$$P^{(\varepsilon)}_{XX}(A,B) = \int_{F_X(A)} du \int_{F_X(B)} dv \; k^{(\varepsilon)}(u,v)$$

where:

$$k^{(\varepsilon)}(u,v) = \frac{\partial^2 K^{(\varepsilon)}(u,v)}{\partial u \partial v}$$

With another change of variables

$$z_1 = \Phi^{-1}(u) \quad \text{and} \quad z_2 = \Phi^{-1}(v)$$

the extended probability can also be written by:

$$P_{XX}^{(\varepsilon)}(A,B)=\int_{\Phi^{-1}(F_X(A))} dz_1 \int_{\Phi^{-1}(F_X(B))} dz_2 \ \varphi_\varepsilon(z_1,z_2)$$

Application of the use of extended likelihood ratio in the case of height measurements of people from recorded videotape is reported in chapter 7.
In order to summarize the obtained results, the principal used quantities are reported in the following table.

Table 5.3.1. *Summary of the used quantities.*

1-D spaces	R		$[0,1]$		Ω
variables	z		u		s
distributions		$\Phi \rightarrow$		$\leftarrow F_X$	
densities		$\phi \rightarrow$		$\leftarrow p_X$	
differentials	dz	$=$	$d\Phi^{-1}(u)$	$=$	$d\Phi^{-1}(F_X(s))$
	$d\Phi(z)$	$=$	du	$=$	$dF_X(s)$
	$dF_X^{-1}(\Phi(z))$	$=$	$dF_X^{-1}(u)$	$=$	ds
2-D spaces	R^2		$[0,1]^2$		Ω^2
distributions	$\Phi^{(\varepsilon)}(z_1,z_2)$		$K^{(\varepsilon)}(u,v)$		$H_{XX}^{(\varepsilon)}(s,t)$
densities	$\varphi_\varepsilon(z_1,z_2)$		$k^{(\varepsilon)}(u,v)$		$h_{XX}^{(\varepsilon)}(s,t)$
marginals	$\phi(z)$		1		$p_X(s)$
$\lim_{\varepsilon\rightarrow 0}$	$\phi(z_1)\delta(z_1-z_2)$		$\delta(u-v)$		$p_X(s)\delta(s-t)$

5.4. CONCENTRATION PROPERTIES OF THE COPULA

In order to analyze the properties of the copula, a parameter λ is introduced in evaluating how much the density is concentrated along the diagonal $u=v$. The choice of this parameter, which is a function of ε, is the following:

$$\lambda(\varepsilon)=\int_0^1 du \int_0^1 dv \ [u-v]^2 \ k^{(\varepsilon)}(u,v)$$

According to the change of variables of the previous section, the parameter can also be written:

$$\lambda(\varepsilon)=\int_R dz_1 \int_R dz_2 \ [\Phi(z_1)-\Phi(z_2)]^2 \ \varphi_\varepsilon(z_1,z_2)$$

The choice of the bivariate Gaussian distribution allows the explicit computation of this function (section 5.5), with $\varepsilon \in (0,1)$:

$$\lambda(\varepsilon) = \frac{1}{\pi} \arctan \frac{\sqrt{(1+\varepsilon)(3-\varepsilon)}}{1-\varepsilon} - \frac{1}{3}$$

Obviously,

$$\lim_{\varepsilon \to 0} \lambda(\varepsilon) = 0$$

and more precisely:

$$\lambda(\varepsilon) = \frac{\sqrt{3}}{3\pi} \left[\varepsilon - \frac{1}{6}\varepsilon^2 + o(\varepsilon^2) \right]$$

In the case of independent variables, i.e. for $\varepsilon = 1$, trivially the density of the copula is $k^{(\varepsilon)}(u,v) = 1$. Hence, $\lambda(1)$ can be easily computed according to its definition and the result is $\lambda(1) = \frac{1}{6}$. The same result can be also obtained from its explicit formula: $\lim_{\varepsilon \to 1} \lambda(\varepsilon) = \frac{1}{6}$.

5.5. CALCULATION OF $\lambda(\varepsilon)$

In the following, some integrals known in literature are used, and in particular formulas [3.322.2] and [6.285.1] of reference [7]. For sake of linearity, the two integrals are here reported.

$$\int_0^{+\infty} dx \; \exp\left[-\frac{x^2}{4\beta} - \gamma x \right] = \sqrt{\pi\beta} \; \exp(\beta\gamma^2) \left[1 - \Phi_0(\gamma\sqrt{\beta}) \right] \quad , \quad \mathrm{Re}\, \beta > 0$$

$$\int_0^{+\infty} dx \; [1 - \Phi_0(x)] \; e^{-\mu^2 x^2} = \frac{\arctan \mu}{\mu\sqrt{\pi}} \quad , \quad \mathrm{Re}\, \mu > 0$$

where:

$$\Phi_0(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-s^2} ds$$

Note that it is easy to show that:

$$\int_0^{+\infty} dx \; [1 - \Phi_0(-x)] \; e^{-\mu^2 x^2} = \frac{1}{\mu\sqrt{\pi}} [\pi - \arctan \mu]$$

using the fact that $\Phi_0(x) = -\Phi_0(-x)$ and the Gaussian integral $\int_0^{+\infty} dx \; e^{-\mu^2 x^2} = \frac{\sqrt{\pi}}{2\mu}$. Now, let A be a real symmetric 2×2-matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}$$

with $a_{11} > 0$ and $\det A > 0$. Then:

$$\int_0^{+\infty} dz_1 \int_0^{+\infty} dz_2 \ \exp\left[-\frac{1}{2}\mathbf{z}A\mathbf{z}\right] = \begin{cases} \frac{1}{\det^{1/2} A} \arctan\left[\frac{\det^{1/2} A}{a_{12}}\right] & , \quad a_{12} > 0 \\ \frac{1}{\det^{1/2} A}\left(\pi - \arctan\left[\frac{\det^{1/2} A}{|a_{12}|}\right]\right) & , \quad a_{12} < 0 \end{cases}$$

Actually, applying the known integrals, the left hand side integrated with respect to z_1 is:

$$\sqrt{\frac{\pi}{2a_{11}}}\int_0^{+\infty} dz_2 \ \left[1 - \Phi_0(\frac{a_{12}z_2}{\sqrt{2a_{11}}})\right] \ \exp\left[-\frac{z_2^2}{2a_{11}}\det A\right]$$

and in the two cases $a_{12} > 0$ and $a_{12} < 0$ the the results directly follow.
Now, in order to calculate explicitly $\lambda(\varepsilon)$, a remark is preliminary made:

$$\begin{aligned}\lambda(\varepsilon) &= \int_R dz_1 \int_R dz_2 \ [\Phi(z_1) - \Phi(z_2)]^2 \ \varphi_\varepsilon(z_1, z_2) \\ &= \frac{2}{3} - 2\alpha(\varepsilon)\end{aligned}$$

where the fact that:

$$\int_R dz_1 \Phi^2(z_1) \int_R dz_2 \varphi_\varepsilon(z_1, z_2) = \int_R dz_1 \Phi^2(z_1)\phi(z_1) = \left.\frac{\Phi^3(z_1)}{3}\right|_0^1 = \frac{1}{3}$$

and its symmetric relation are used; moreover, by definition:

$$\alpha(\varepsilon) = \int_R dz_1 \int_R dz_2 \ \Phi(z_1)\Phi(z_2) \ \varphi_\varepsilon(z_1, z_2)$$

Using the integral representation of Φ, the parameter $\alpha(\varepsilon)$ can be written:

$$\alpha(\varepsilon) = \int_{R^2} d\mathbf{z} \int_{\mathbf{s}\leq\mathbf{z}} d\mathbf{s} \ \varphi_1(\mathbf{s})\varphi_\varepsilon(\mathbf{z})$$

and, making a variable substitution with unitary Jacobian:

$$\mathbf{z} = \mathbf{w} + \frac{\mathbf{t}}{2} \quad \text{and} \quad \mathbf{s} = \mathbf{w} - \frac{\mathbf{t}}{2}$$

the subspace of R^4, $\{\mathbf{z}, \mathbf{s} | \mathbf{s} \leq \mathbf{z}\}$, becomes $\{\mathbf{w}, \mathbf{t} | \mathbf{t} \geq \mathbf{0}\}$, and so:

$$\alpha(\varepsilon) = \int_{R^2} d\mathbf{w} \int_{(R^+)^2} d\mathbf{t} \ \varphi_1(\mathbf{w} - \frac{\mathbf{t}}{2})\varphi_\varepsilon(\mathbf{w} + \frac{\mathbf{t}}{2})$$

Performing the Gaussian integral with respect to the variables $\mathbf{w}$:

$$\alpha(\varepsilon) = \frac{1}{2\pi \cdot \det^{-1/2}[1 + C_\varepsilon]} \int_{(R^+)^2} d\mathbf{t} \ \exp\left[-\frac{1}{2}\mathbf{t}[1 + C_\varepsilon]^{-1}\mathbf{t}\right]$$

where the relation:

$$[1 + C_\varepsilon^{-1}] - [1 - C_\varepsilon^{-1}] \cdot [1 + C_\varepsilon^{-1}]^{-1} \cdot [1 - C_\varepsilon^{-1}] = 4 \cdot [1 + C_\varepsilon]^{-1}$$

has been used.
Finally, explicit calculation of $\alpha(\varepsilon)$ follows from the fact that the last formula can be estimated by the previous results, in the case of $a_{12} < 0$. So:

$$\alpha(\varepsilon) = \frac{1}{2\pi}\left[\pi - \arctan\frac{\sqrt{(1 + \varepsilon)(3 - \varepsilon)}}{1 - \varepsilon}\right]$$

The conclusion is:

$$\lambda(\varepsilon) = \frac{1}{\pi}\arctan\frac{\sqrt{(1 + \varepsilon)(3 - \varepsilon)}}{1 - \varepsilon} - \frac{1}{3}$$

REFERENCES

[1]. Karlin Samuel, Taylor Howard M., *A first course in stochastic processes*, Academic Press (1975)

[2]. Rohatgi Vijay K., *An introduction to probability theory and mathematical statistics*, John Wiley & Sons (1976)

[3]. Taroni Franco, Mangin Patrice, *L'interprétation de la preuve scientifique: les juristes, les scientifiques et les probabilités*, Médicine & Droit, 30, 6-15 (1998)

[4]. Nelsen Roger B., *An introduction to copulas*, Lecture notes in statistics, 139, Springer-Verlag Telos (1999)

[5]. Curran James M., Hicks Tacha N., Buckleton Jhon S., *Forensic interpretation of glass evidence*, CRC Press (2000)

[6]. `http://www.R-project.org`

[7]. Gradshteyn Izrail S., Ryzhik Iosif M., *Table of integrals, series, and products*, Fifth edition, Alan Jeffrey Editor, Academic Press (1994)

SECTION II

NON-STANDARD APPLICATIONS

6

GUNSHOT RESIDUES CONTAMINATION

The detection of gunshot residue (GSR) particles on suspect's body or objects have always represented an important information in investigation and a determinant evidence in Court in order to determine if the suspect has shot firearms. However, the typical defender's objection in the Court in the case of positive results in finding GSRs is the possibility of an accidental contamination, in particular since the operators who collect particles are often police officers. Although the collection kits normally contain a series of precautions in order to avoid the risk of accidental contamination of a suspect, such possibility cannot be excluded.

In order to evaluate the likelihood ratio of accidental contamination, two Poisson distributions have been taken into account: the parameter λ of the first one coincides with the mean number of GSRs that can be found on a firearm shooter, while the parameter μ of the second one is the mean number of GSRs that can be found on a non-shooter. In this scenario, the likelihood ratio of finding n gunshot residues can be easily calculated. The evaluation of the two parameters is performed by using two sets of data: "exclusive" lead-antimony-barium gunshot residues have been detected on a population of 31 police officers after firearm practice, and on a population of 81 police officers who had declared not to have handled firearms since almost one month. The data show that the detection of two or more GSRs normally favours the hypothesis that the suspect has shot firearms.

6.1. STATISTICAL FRAMEWORK

In order to evaluate the likelihood ratio about the possibility of GSR accidental contamination of a suspect, the following statistical framework is delineated.

Let C and $\Omega - C$ denote the prosecutor's and the defender's hypotheses, respectively: the hypothesis C considers that the suspect has shot a firearm, while the hypothesis $\Omega - C$ supposes that the suspect has not:

(a). C: the suspect has shot a firearm;

(b). $\Omega - C$: the suspect has not shot a firearm.

Now, for $n \geq 0$, let E_n be the evidence that n GSRs have been detected on the suspect:

(c). E_n: n GSRs have been detected on the suspect;

and let $P(E_n|C)$ and $P(E_n|\Omega - C)$ denote the conditional probabilities of the event E_n in the prosecutor's and in the defender's hypotheses, respectively. In other words, $P(E_n|C)$ represents the probability of detecting n GSRs in the case the suspect has shot a firearm, while $P(E_n|\Omega - C)$ represents the probability of finding n GSRs in the case of accidental contamination. According to the Bayesian approach, the aim of the forensic scientist lies in the calculation of the likelihood ratio $Lr(n)$, which is defined by the ratio of the two conditional probabilities:

$$Lr(n) = \frac{P(E_n|C)}{P(E_n|\Omega - C)}$$

If $Lr(n) > 1$, the detection of n GSRs favours the prosecutor's hypothesis C, and the greater the value of the likelihood ratio is, the more determinant the evidence E_n is. On the contrary, if $Lr(n) < 1$, the defender's hypothesis $\Omega - C$ is stronger than the prosecutor's one, and $Lr(n) \ll 1$ means a very high probability of accidental contamination. Finally, $Lr(n) = 1$ implies the absolutely unconcern of the evidence about the two hypotheses.

In order to evaluate the two probabilities, the GSR presence on the samples has been considered as a random event, and so the proposed distribution in estimating the count of their number is the Poisson one.

In this scenario, let λ be the mean number of detected GSRs on a shooting person; so, the probability distribution of finding n GSRs in the prosecutor's hypothesis is:

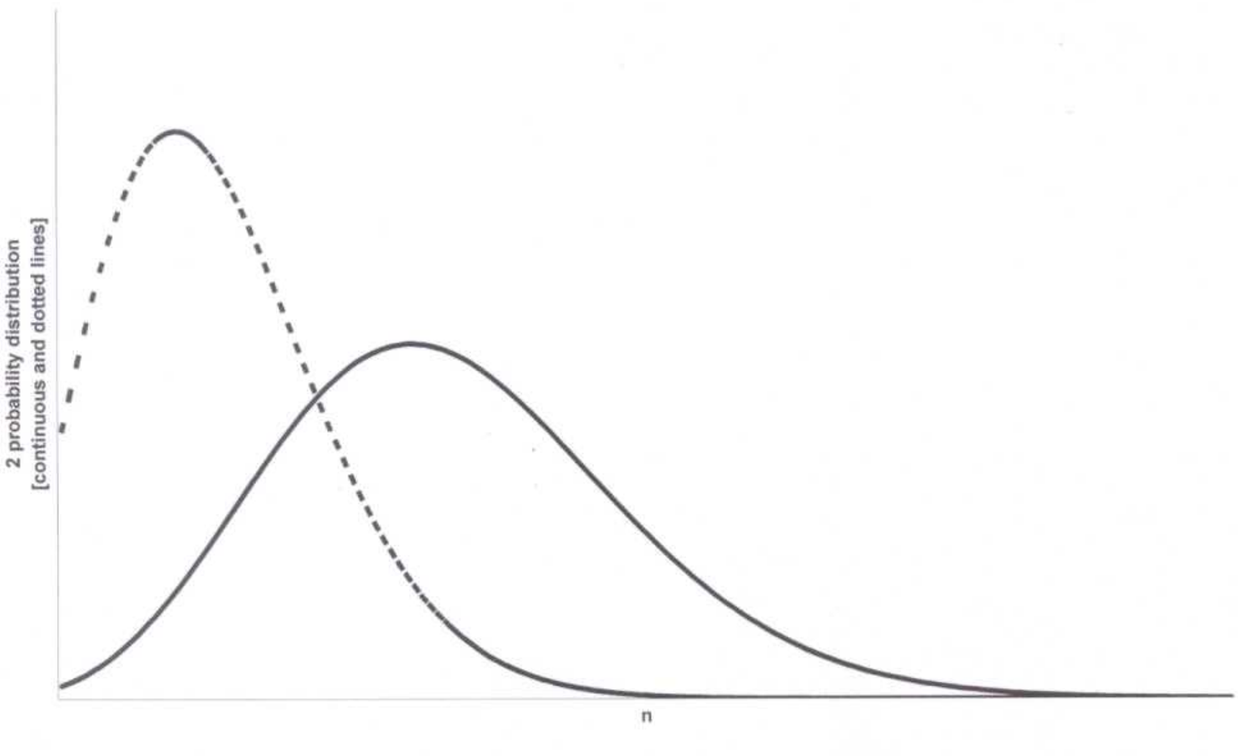
$$P(E_n|C) = e^{-\lambda} \cdot \frac{\lambda^n}{n!}$$

Analogously, if μ is the mean number of detected GSRs on a non-shooting person, the probability distribution of finding n GSRs by accidental contamination is:

$$P(E_n|\Omega - C) = e^{-\mu} \cdot \frac{\mu^n}{n!}$$

Obviously, normally λ is expected to be greater than μ.

Figure 6.1.1. *Graphics of two Poisson distributions with parameters λ [continuous line] and μ [dotted line], with $\lambda > \mu$.*



According to these statements, the likelihood ratio $Lr(n)$ about the event E_n in the conditional hypotheses C and $\Omega - C$, can be calculated, and it reduces to:

$$Lr(n) = e^{\mu-\lambda} \cdot \left(\frac{\lambda}{\mu}\right)^n$$

Since λ is supposed to be greater than μ, the likelihood ratio $Lr(n)$ is an increasing exponential function of n. This means that the greatest is the number of the detected GSRs, the smallest is the chance of accidental contamination. Moreover:

$$Lr(n) > 1 \quad \text{for} \quad n > \frac{\lambda - \mu}{\ln \lambda - \ln \mu}$$

6.2. DATA

In order to evaluate the two parameters introduced in the statistical framework, two different series of particle collection have been performed by means of the standard kits normally used in forensic laboratories. First, particles have been collected on the hands of 31 police officers who had shot 10 shots by pistol Beretta mod. 85F, cal. 9×17 mm. Collections have been taken at different time since the shot (after $t = 2$, 3, 4, 5, 6, 8, and 10 hours). Secondly, particles have been collected on the hands of 81 police officers declaring not to have handled firearms since almost one month, but normally living and working in the same spaces (rooms, cars, canteen, etc.) as the shooting police officers. All the samples have been carbon coated to increase the electrical conductivity, and analyzed by scanning electron microscopy/energy-dispersive spectrometry (SEM/EDS) by automated and manual methods. The operating conditions for SEM/EDS are reported in the following table.

Table 6.2.1. *Operating conditions for SEM/EDS.*

Condition	Setting
accelerating voltage	25 kV
working distance	25 mm
specimen tilt	0
magnification	375 $\times$
specimen current	1.8 nA
EDX acquisition energy	1-20 keV

For sake of simplicity, only "exclusive" lead-antimony-barium particles have been taken in consideration as GSRs in this paper, and the results of the two sets of data are listed in the following tables. The mean numbers $\lambda(t)$ and μ of detected GSRs per collection are also calculated and reported in the same tables.

Table 6.2.2. *Results concerning the collection of GSRs on 31 police officers after shooting 10 shots with Beretta mod. 85F, cal.* 9×17 *mm.*

collection condition	**nr. of samples**	**nr. of GSRs**	$\lambda(t)$
after 2 hours	6	97	16.17
after 3 hours	5	49	9.80
after 4 hours	5	43	8.60
after 5 hours	6	39	6.50
after 6 hours	4	21	5.25
after 8 hours	3	11	3.67
after 10 hours	2	5	2.50
TOTAL	31	265	

Table 6.2.3. *Results concerning the collection on 81 police officers declaring not to have handled firearms since almost one month.*

collection condition	**nr. of samples**	**nr. of GSRs**
police officers usually working in cars	20	1
police officers after having handcuffed	12	0
police officers usually working in police station	37	0
police officers usually working in civilian	12	0
TOTAL	81	1
MEAN		$\mu = 0.0123$

The calculated parameter $\lambda(t)$ varies from the value of 16.2 GSRs after 2 hours to the value of 2.5 GSRs after 10 hours. Moreover, since only one GSR has been detected on 81 samples, the mean number of GSRs on non-shooting people can be evaluated $\mu = 0.012$. This data can be compared with those present in literature.

6.3. RESULTS

The parameters calculated from the experimental data allow the authors to evaluate the likelihood ratio, and the values for $n = 0 \dots 10$ are listed in the following table, where the base-10 logarithm of values of the likelihood ratio $Lr(n)$ is reported.

Table 6.3.1. *The base-10 logarithm of the likelihood ratio $Lr(n)$ calculated from the experimental data, for $n = 0 \dots 10$. The negative values are reported in italic characters.*

n	**2 hr**	**3 hr**	**4 hr**	**5 hr**	**6 hr**	**8 hr**	**10 hr**
0	*-7.0*	*-4.3*	*-3.7*	*-2.8*	*-2.3*	*-1.6*	*-1.1*
1	*-3.9*	*-1.4*	*-0.9*	*-0.1*	0.4	0.9	1.2
2	*-0.8*	1.5	2.0	2.6	3.0	3.4	3.5
3	2.3	4.4	4.8	5.3	5.6	5.8	5.8
4	5.5	7.3	7.6	8.1	8.2	8.3	8.1
5	8.6	10.2	10.5	10.8	10.9	10.8	10.5
6	11.7	13.1	13.3	13.5	13.5	13.2	12.8
7	14.8	16.0	16.2	16.2	16.1	15.7	15.1
8	17.9	18.9	19.0	19.0	18.8	18.2	17.4
9	21.0	21.8	21.9	21.7	21.4	20.7	19.7
10	24.2	24.7	24.7	24.4	24.0	23.1	22.0

According to these data, the likelihood ratio is greater than 1 (which implies positive values for $\log_{10} Lr(n)$) - i.e. in the hypothesis that the suspect has shot a firearm the evidence is stronger than in the hypothesis that he has not - in the case of finding two or more GSRs after more than 2 hours. Also the detection of one GSR after 6 or more hours favours the prosecutor's hypothesis. On the contrary, in the other cases, accidental contamination is more probable. Moreover, the results show that the detection of more than three GSRs implies a very high value of the likelihood ratio, and so the probability of accidental contamination in these cases can be considered negligible.

Obviously, in real cases the forensic expert has to take into account a lot of other features: the possibility the suspect has washed hands after the shot, atmospheric conditions (wind, rain, etc.), number of shots, collection efficiency of the samples, and so on.

Finally, also "characteristic" particles, together with "exclusive" lead-antimony-barium ones, could be considered in evaluating the parameters λ and μ, even by introducing a weighted mean. However, the validity of the proposed statistical framework persists, the calculation of the values of likelihood ratio and their interpretation being liable to modification.

REFERENCES

[1]. Aitken Colin G.G., *Statistics and the evaluation of evidence for forensic scientists*, John Wiley & Sons (1997)

[2]. Wolten G.M., Nesbitt R.S., Calloway A.R., Loper G.L., Jones P.F., *Equipment system improvement program Final report on particle analysis for gunshot residue detection*, Report ATR-77(7915)-3, The Aerospace Corporation, El Segundo, CA (1977)

[3]. Basu Samarendra, *Formation of gunshot residues*, J Forensic Sci 27, 72-91 (1982)

[4]. Wallace J.S., McQuillan J., *Discharge residues from cartridge-operated industrial tools*, J Forensic Sci 24, 5, 495-508 (1984)

[5]. Schwoeble A.J., Exline David L., *Current methods in forensic gunshot residue analysis*, CRC Press (2000)

[6]. *Standard guide for gunshot residue analysis by scanning electron microscopy/energy-dispersive spectrometry*, ASTM standards, Designation E 1588-95 (Re-approved 2001)

[7]. Gialamas Dean M., Rhodes Edward F., Sugarman Loren A., *Officers, their weapons and their hands: an empirical study of GSR on the hands of non-shooting police officers*, J Forensic Sci 40 (6), 1086-1089 (1995)

[8]. Jager L., Larsson M., *The value of primer gunshot residue (p-GSR) evidence: a pilot study regarding p-GSR contamination*, Scanning 26 (2), 67-68 (2004)

[9]. Cardinetti Bruno, Ciampini Claudio, D'Onofrio Carlo, Orlando Giovanni, Gravina Luciano, Ferrari Francesco, Di Tullio Donatello, Torresi Luca, *X-ray mapping technique: a preliminary study in discriminating gunshot residue particles from aggregates of environmental occupational origin*, Forensic Science International, 143 (1), 1-19 (2004)

[10]. Romolo Francesco S., Margot Pierre, *Identification of gunshot residue: a critical review*, Forensic Science International, 119, 195-211 (2001)

7

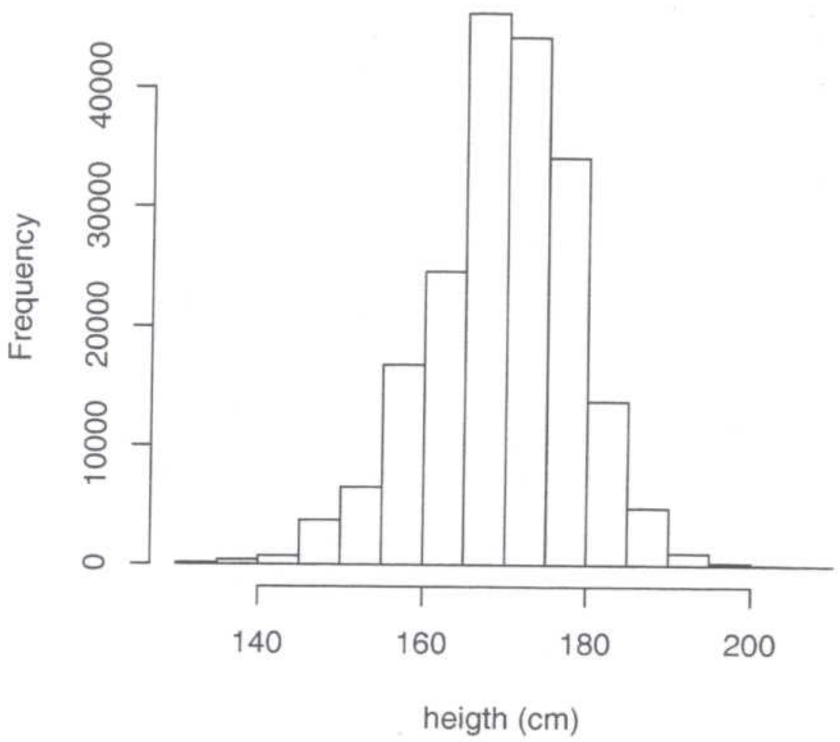
HEIGHT MEASUREMENTS

In order to explain the approach based on extended likelihood ratio, application of this method in the case of height of persons measured from a recorded videotape. The computation of the extended probability is performed according to Italian Carabinieri database. Disjoint but very close intervals are taken as prosecutor's hypothesis C and evidence E.

7.1. DATA

Let X denote the random variable which models the height of people in the population in the probability space $(\{\Omega, \mathcal{E}, P\}$, and let P_X indicate the relative probability measure.
The probability is computed by taking in account the real value of the database of about 200,000 people filed in the police archive of Italian Carabinieri.
The histogram of height measurements in centimeters is reported in the following figure: the measured mean value is $\mu = 170.73$, while the standard deviation is $\sigma = 8.86$.

Figure 7.1.1. *Histogram of the heights extracted from the database of about 200,000 people filed in police archive of Italian Carabinieri. Mean value:* $\mu = 170.73$*; standard deviation:* $\sigma = 8.86$.



7.2 EVALUATION OF THE EXTENDED LIKELIHOOD RATIO

The proposed approach of the extended likelihood ratio is here applied to a case-work in which the measurements of the height of the prosecutor's hypothesis is the interval $C = [x_1, x_2]$ and the evidence is the interval $E = [y_1, y_2]$.

These two sets define a rectangle $(C, E) \subseteq R^2$. It is well known that the probability of this rectangle with respect to $P_{XX}^{(\varepsilon)}$ can be computed using the associated distribution function $H_{XX}^{(\varepsilon)}$ according to the formula:

$$P_{XX}^{(\varepsilon)}(C, E) = H_{XX}^{(\varepsilon)}(x_2, y_2) + H_{XX}^{(\varepsilon)}(x_1, y_1) - H_{XX}^{(\varepsilon)}(x_1, y_2) - H_{XX}^{(\varepsilon)}(x_2, y_1)$$

The following version of formula for evaluation of $P_{XX}^{(\varepsilon)}(C, E)$ can be used:

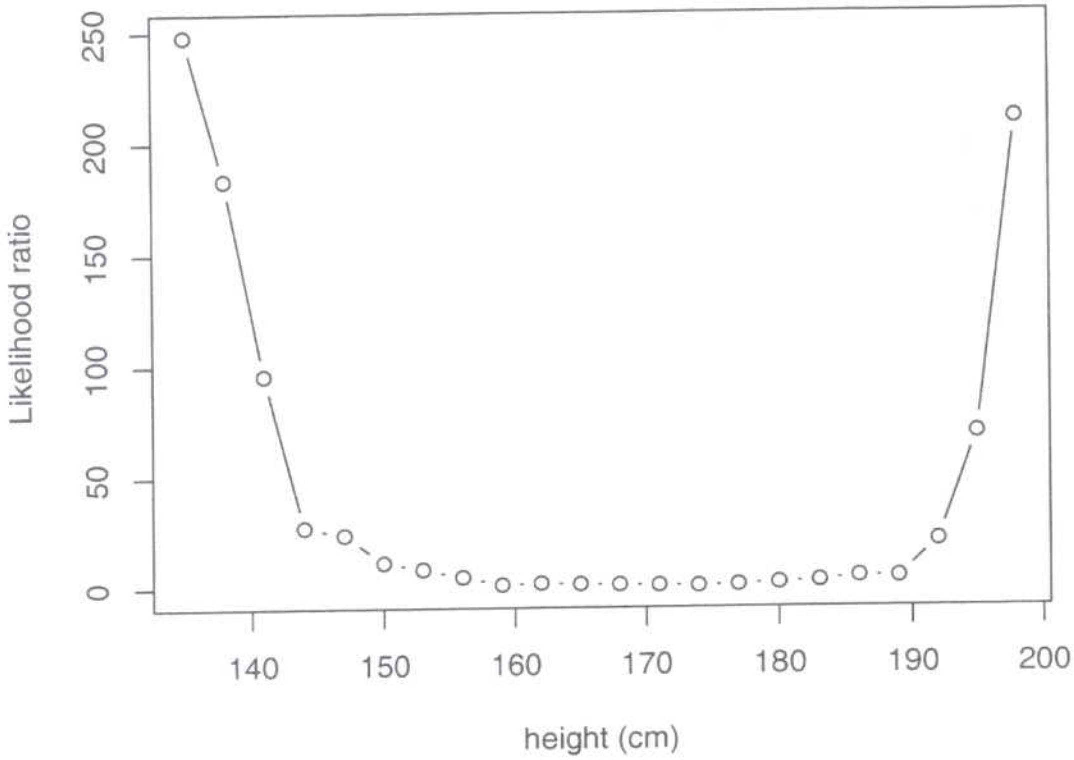
$$P_{XX}^{(\varepsilon)}(C, E) = \int_{\Phi^{-1}([F_X(x_1), F_X(x_2)])} dz_1 \int_{\Phi^{-1}([F_X(y_1), F_X(y_2)])} dz_2 \; \varphi_\varepsilon(z_1, z_2)$$

For example, from the vision of recorded videotape, the error in measurements of height is 4 centimeters, and the the intervals of the evidence E and of the prosecutor's hypothesis C is 1 centimeter, e.g. $E = [x - 5, x]$ and $C = [x + 1, x + 6]$. Note that E and C are disjoint but very close intervals.

Choosing the correlation parameter $\varepsilon = 0.05$, the extended likelihood ratio $Lr^{(\varepsilon)}(E; C)$ is calculated numerically, using the free statistical software `R`.

This choice defines a function of x with the following properties: it is always positive, its value is about 1 if x is near the average of the heights, and it is very large if x is in the tails of the distribution, as expected. The plot of this function is shown in the following figure.

Figure 7.2.1. *Extended likelihood ratio as function of the height x for $\varepsilon = 0.05$, for the disjoint but very close intervals $E = [x - 5, x]$ and $C = [x + 1, x + 6]$.*



This way of evaluation of the likelihood ratio allows to quantify the negative conclusion cases. Moreover, the quantification takes into account the statistical distribution of the values in the population. If the value is about 1 (unconcern) if x is near the average of the heights; on the contrary, in the tails of the distribution the value of the extended likelihood ratio grows up, as expected, and an important contribution in evaluation of odds has to be considered.
The method can be applied in all the branches of forensic science, and in particular to continuous characteristics.

REFERENCES

[1]. `http://www.R-project.org`